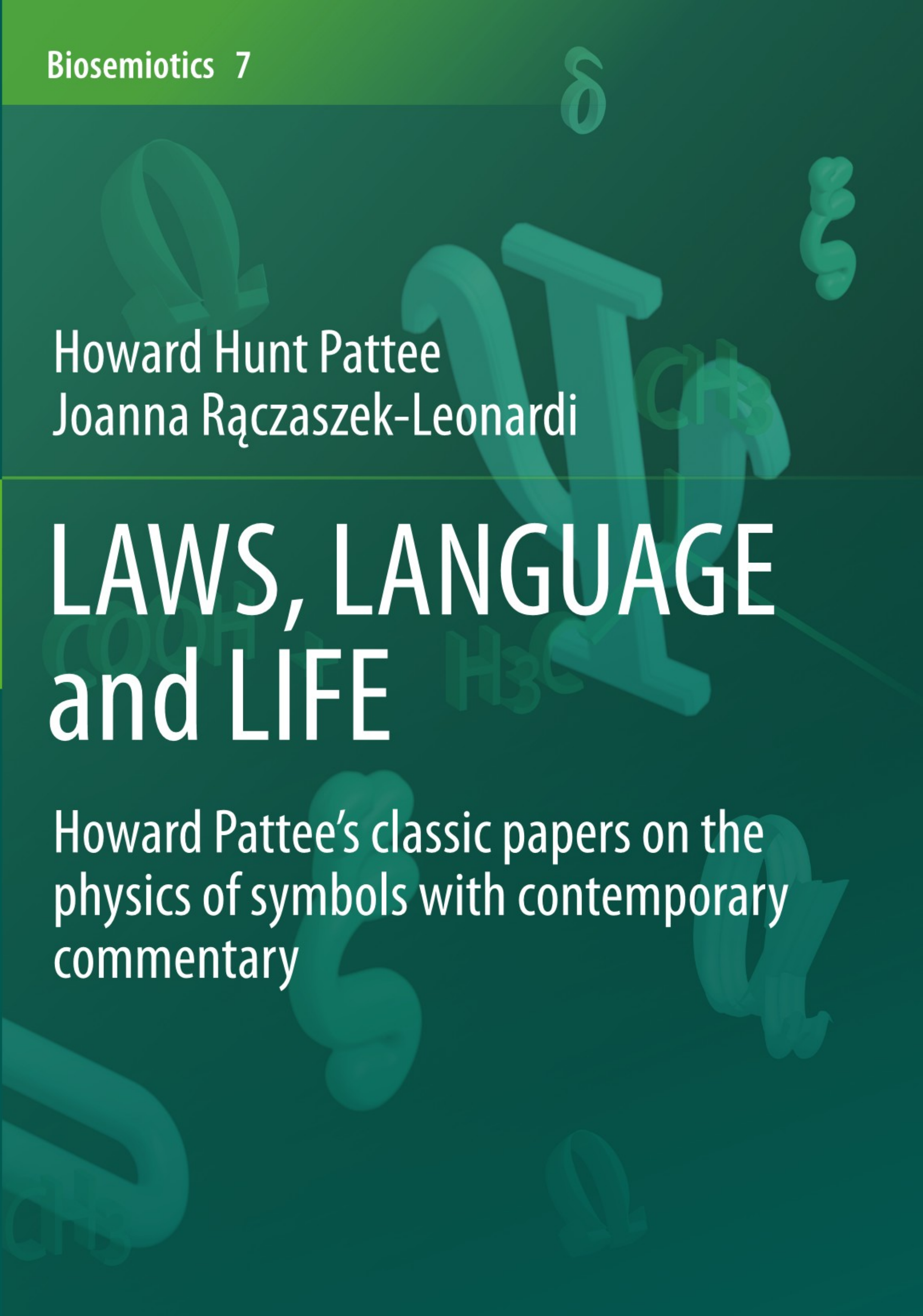

Biosemiotics 7

Howard Hunt Pattee
Joanna Rączaszek-Leonardi

LAWS, LANGUAGE and LIFE

Howard Pattee's classic papers on the physics of symbols with contemporary commentary

Springer

LAWS, LANGUAGE and LIFE

BIOSEMIOTICS

VOLUME 7

Aims and Scope of the Series

Combining research approaches from biology, philosophy and linguistics, the emerging field of biosemiotics proposes that animals, plants and single cells all engage in semiosis – the conversion of physical signals into conventional signs. This has important implications and applications for issues ranging from natural selection to animal behaviour and human psychology, leaving biosemiotics at the cutting edge of the research on the fundamentals of life.

The Springer book series *Biosemiotics* draws together contributions from leading players in international biosemiotics, producing an unparalleled series that will appeal to all those interested in the origins and evolution of life, including molecular and evolutionary biologists, ecologists, anthropologists, psychologists, philosophers and historians of science, linguists, semioticians and researchers in artificial life, information theory and communication technology.

For further volumes:
http://www.springer.com/series/7710

Howard Hunt Pattee • Joanna Rączaszek-Leonardi

LAWS, LANGUAGE and LIFE

Howard Pattee's classic papers on the physics of symbols with contemporary commentary

Springer

Howard Hunt Pattee
Binghamton University
NY, USA

Joanna Rączaszek-Leonardi
University of Warsaw
Poland

ISSN 1875-4651 ISSN 1875-466X (electronic)
ISBN 978-94-007-5160-6 ISBN 978-94-007-5161-3 (eBook)
DOI 10.1007/978-94-007-5161-3
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2012944987

There are instances where we have been unable to trace or contact the copyright holder. If notified the publisher will be pleased to rectify any errors or omissions at the earliest opportunity.

Acknowledgements: work on this book was supported by Marie Curie ERG Grant PERG06-GA-2009-256587 and BST 161551/2011 to the second author.

© Springer Science+Business Media Dordrecht 2012
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface: a Conversation

Joanna Rączaszek-Leonardi: Finally, a book with selected reprints of your work! We've needed it for a long time. I know that your perspective on the reasons for reprinting these papers is different than my own, but the first stimulus for me was the obvious relevance of your papers to the problems in modern cognitive science that are increasingly harder to sweep under the carpet. Problems so fundamental, that many—even the mainstream—researchers feel frustration: During the *Decade of the Brain* (1990s) numerous brain imaging techniques were developed and perfected; the next two decades witnessed an unprecedented quantity of experimental research on human brain and cognition—from the molecular level of single neuron functioning to complex reasoning in social groups. Yet the relation between this immense collection of facts about the biochemical and physiological properties of the brain and our conscious, culturally infused, experience is still largely a mystery. I think your work, even though it concerns the level that appears remote from the functioning of complex organization of human brain and cognition, cuts right through to the reasons for this chasm. So this was the first motivation. The second was a simple annoyance at not being able to find your works, especially those published in the 1960s to 1980s.

Howard Pattee: Your interest in reprinting this selection of my papers, which were written over a period of 45 years, made me reconsider the general nature of my subject matter. My papers were published first as biophysics and then as theoretical biology. These were popular fields when I began writing. Later they were motivated by origin of life and artificial life studies, and the interest spread to other areas, as the titles of my papers indicate. Now they are being reprinted here in another area called *biosemiotics*. Today, I would say that the deeper motivation for all these papers, as I explain in my introductory historical commentary, belongs to the branch of philosophy called epistemology. The central issue of traditional epistemology is how the knowledge inside human brains corresponds to what exists outside our brains, but I saw this subject-object problem at a much simpler level beginning at the origin of life where, instead of simply a passive correspondence problem, it becomes a problem of how genetic symbols construct and control the replicating cell. The recognition that symbolic control is the basis of all life is now the view that defines *biosemiotics*.

Joanna: This is probably how “basic” we have to go to rethink our notions of what cognition is. If one thinks about cognitive systems broadly, as systems retaining information to adapt to their environments, then biosemiotics and cognitive science have a much broader range of problems in common than it is usually assumed.

Howard: My first papers are about the classic problem of the physical basis of life—not in terms of abiogenic chemistry, but as a conceptual problem of where symbolic function emerges in the context of physical laws—laws that we express in mathematical symbols, but laws that do not control symbolic behavior, and say nothing about it. This conceptual problem of how specific but arbitrary symbol systems interact with inexorable physical laws arises at all evolutionary levels, from the genetic language to human language, logic, mathematics, and computation.

Actually, I see the situation in biology as having a parallel to the one you described for the cognitive sciences. Over the years during which I wrote these papers, there have been enormous increases in knowledge in genetics and molecular biology, as well as in the newer cognitive sciences. In spite of this increase in knowledge, these advances have rarely clarified the epistemological problems of the separation of subject and object, and the relation of symbols to matter—indeed, in the case of quantum theory, I think advances have made the problem even more obscure. Automated instruments, computers, and vast amounts of memory storage have produced far more scientific data in the last 50 years than in all of previous history, but collecting more data is not likely to help. In my view, epistemology is about what it means for individual agents, from cells to humans, to *make sense* of their data. That is the underlying problem in these papers. All the subjects I discuss in the papers in this collection developed from this epistemological problem of understanding how subjective function and meaning arise from the objective stream of events.

Joanna: Thus it is studying the subject-object relation at the simplest levels that may offer a step in answering *how* physical events become meaningful for higher organisms. This question may take many forms. For biologists, the question is, quoting the title of your second reprinted paper “How does a molecule become a message”; for cognitive scientists, psychologists and linguists, it may take a form of asking how natural language symbols relate to the dynamics in which they are immersed and from which they arose. The relation between these questions at such vastly different evolutionary levels is far from simple.

Howard: Even in the context of classical physics, the origin of symbols is an obscure problem. In quantum mechanics, what is called the *measurement problem* is even more obscure. It arises when a physical interaction of a measuring device with a quantum system results in a classical record. This record has the specificity and arbitrariness characteristic of what is called a symbol. Neither classical nor quantum laws can determine when a measurement occurs. The additional problem with quantum mechanics is that it cannot describe the classical symbolic result. When I began graduate study in physics in the 1940s, many prominent physicists (e.g., Bohr, Schrödinger, Heisenberg, Pauli, Delbrück, Wigner) doubted that life could be adequately described by quantum laws. In a sense that I explain in these papers, I think that they were right.

However, by the 1960s, after the discovery of the DNA double helix and the genetic code, almost all these doubts about the adequacy of quantum laws were simply ignored, and molecular biology took over, with classical chemistry apparently providing adequate models for biologists. But the measurement problem and the problem of when quantum models can be replaced by classical models is still a foundational issue for physicists. What physicists agree on is that measurement and observation, in both classical and quantum models, require a clear distinction between the objective events and subjective records of events. This is not an ontological distinction, but follows from the necessity of what I call an *epistemic cut*—a concept that in many ways ties together all these papers.

Joanna: The concept of epistemic cut, which involves complementarity of a discrete symbolic and continuous dynamic mode, was missing also from the approaches to cognition that dominated over the last 50 years. Your papers allow us to step back to the period of the mid last century, when cognitive sciences were born in the excitement of the postwar technological developments, and at the same time disappointment with the then dominating (at least in the USA) behaviorist framework. There were probably many ways in which to oppose behaviorism and recognize that inner states and processes are important in the explanation of human behavior. But for some reasons, only two models established themselves as independent schools: the information processing paradigm, searching for processes compatible with Artificial Intelligence, based on computer simulations (and largely funded for that reason); and the opposing views of ecological psychology (founded by J. J. Gibson).

Your work shows that already in this time, at the beginning, a third way existed. Your arguments are based on fundamental physics, but their philosophical basis appears compatible with certain schools in philosophy, such as the phenomenological approach. However this does not mean that they are a threat to more analytically-based approaches in cognitive sciences. By showing the indispensability of symbols and their role in a dynamical biological organization, this view has a potential for bridging the complementary symbolic and dynamic approaches to cognition, as well as specifying the role of the observer-researcher in the discovery process. As a cognitive scientist I am excited about this perspective of reconciliation. But there is also another, not less important consequence of applying your framework: that of situating the problems of human language and cognition within a broader theory of information in all living systems.

Howard: I found that the interest in reprinting these papers also comes from ex-students and colleagues, as well as from biosemioticians who agree with my view that the origin of life, all of evolution, and all languages exhibit an agent's symbolic control of matter. The choice of papers was influenced by the recommendations of these groups, and by an attempt to cover the diverse fields and audiences for whom they were written. These fields include physics, molecular and developmental biology, evolution, cognitive science, artificial intelligence, artificial life, sociology, semiotics, and linguistics. I have learned something about these other fields mostly by reading their literature and participating in their meetings, but I do not consider myself an expert in any of them. However, at the epistemological level, the questions I address apply to all of them.

Joanna: The diversity of your subjects may be a problem for some readers. Do you have any advice on how the papers should be read? Personally, I was impressed that your papers – which, after all, belong to the domain of physics, theoretical biology and theory of information – can be so readable and instructive for a psychologist or cognitive scientist.

Howard: My introductory commentary explains the personal historical motivations for many of the papers. I think the scientific field for which each paper was written is clear from the title and the references. The papers are presented here in the order they were published. This may be of historical interest, but the papers are not meant to be read together, or in any order. They are self-contained and can be read individually. Perhaps readers interested in the cognitive sciences may find it helpful to first read your *Afterword* that reviews relevant issues in cognitive science, and places some of my papers in that context.

Joanna: I am very happy that the Biosemiotic series Editor and Springer publishing house gave us the opportunity to consider the birth and evolution of this wider framework in its original form and evaluate its usefulness from many perspectives. Physics, biology, and cognitive science have travelled a fascinating path since the publishing of the first papers in this volume, yet the problems posed there are still of utmost importance. I would like to thank Marcello Barbieri and Catherine Cotton, for believing in this project and encouragement, and Ineke Ravensloot for her editorial work. I would like also to thank Scott Kelso, who first introduced me to your work in the context I describe in my Afterword, and who never ceases to force me out of comfortable conceptual equilibria. To Don Favareau I am indebted for his advices and thorough review of my chapter. I also thank Carol Fowler, Riccardo Fusaroli, Stephen Cowley and Joerg Zinken for their valuable comments. But most of all I thank you for being a patient teacher. Working on this book afforded me a great opportunity: to discuss with you at length the problems I see as fundamental in the present cognitive science.

Howard: Unfortunately, I can no longer recall all the teachers, students, and colleagues that contributed to the ideas expressed in these papers. I must add, however, Robert Rosen and Michael Conrad to the scientists mentioned in my introductory history. They both catalyzed and criticized many aspects of my thoughts, beginning nearly 50 years ago and continuing over several decades. Rosen's ideas on hierarchy theory and on the modeling relation had common features with my own largely because many of them were developed during our discussions. Conrad's understanding of evolution, adaptability, and the limitations of computer models are reflected in my papers. My introductory history and commentary has benefitted from the advice of Peter Cariani and from Donald Favareau's editing. Finally, my sincere thanks go to Joanna who initiated the publication of the present volume and who, in her *Afterword*, has extended my early ideas to the more recent areas of the cognitive sciences.

Contents

Part I Commentary and History—H.H. Pattee

Part II The Pattee Papers in Chronological Order

Part III Afterword - Joanna Rączaszek-Leonardi

Part I
Commentary and History—H.H. Pattee

Introduction—What These Papers Are About

> Men ought to know that from nothing else but the brain come joys, delights, laughter and sports, and sorrows, griefs, despondency, and lamentations. And by this, in an especial manner, we acquire wisdom and knowledge, and see and hear, and know what are foul and what are fair, what are bad and what are good, what are sweet, and what unsavory; some we discriminate by habit, and some we perceive by their utility. Hippocrates (~400 BC)

The types of questions I discuss in these papers are entitled to be called classical questions, because in one form or another they have been on philosophers' minds for well over 2000 years. They arise from the three foundational concepts in the book's title. The first concept is *natural law*, by which I mean the inexorable events over which living organisms have *no control*; or as the physicist Wigner expressed it, a lawful event gives "the impression that it could not be otherwise." The second concept is *life*, and its essential characteristic of *individual* organisms with *variable heritable controls* allowing them to generate a world of endless novelties where, as the biologist Dobzhansky says, "nothing makes sense except in the light of evolution." The third concept is *language*, in which I include all those symbol systems that are necessary for life, evolution, and thought, as well as for control, communication, and models of reality. The genetic language, animal languages, natural human languages, mathematics, formal logic, and computer languages are examples.

Always lurking behind our extensive scientific knowledge of laws, language, and life are the classical philosophers' epistemological questions: By what criteria and actions do the concepts in our individual subjective brains conform to the external objective natural forms in the universe? How much is our knowledge of these forms limited by how our senses and brains have evolved? How much do these forms depend on conceptual and linguistic constraints? How effectively do the brain's cognitive binary oppositions, like discrete and continuous, finite and infinite, time dependent and space (sequence) dependent, determinism and chance, describe the objective forms in the universe? And finally, if life and mind arose only from natural law-abiding earth, air, water, and fire, why are life and mind so peculiarly different?

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7,
DOI 10.1007/978-94-007-5161-3_1, © Springer Science+Business Media Dordrecht 2012

These are the types of questions that motivated these papers. What do I say that is new about these old and controversial classical problems? Unlike traditional philosophers and most cognitive scientists, I do not approach these problems at the level of human brains and human language, but at the level of the origin of life and the cell's genetic language where I believe these epistemological questions first arise, and where they can be most clearly understood. In physics we are taught that we have to understand the simplest system before complex systems can be explained. That does not mean I am a reductionist. I believe that physical laws and biological evolution produce emergent novelties that are recognizable but unpredictable. In fact, I begin with the question: When do emergent concepts like *symbol, language, life, function*, and *meaning* first make sense within the context of physical laws—none of which recognize these concepts?

These epistemological questions are not considered popular or cutting-edge issues in spite of their historical persistence. Most sciences continue in their course without explicit concern for them, and active scientists are too busy with their current research to worry about them. I would even advise young scientists, if their aim is to advance their careers, to ignore these questions and get on with their experiments. On the other hand, I think it will be increasingly difficult to avoid these questions, especially in physics, biology, and the cognitive sciences.

Why do I say this? In physics there is strong empirical evidence that within the size and energy ranges that known life exists, quantum laws are fundamental, and that classical laws are only useful approximations or practical simplifications of quantum laws. The conceptual result of quantum behavior is that our images of the elementary forms in nature are not becoming clearer. In fact, under the current views of quantum theory as the fundamental model, the basic forms of nature appear even more mysterious, and in some cases ineffable, like quantum non-locality. Yet in spite of this evidence that quantum theory is fundamental, we still model the molecules of life as classical structures; and all our scientific models are expressed with the strictly classical symbol vehicles that we use for language, mathematics, and computational codes. In fact classical material symbol structures are all that exist to form the languages describing physical laws, including quantum laws. The problem is that using *only* the fundamental quantum laws there is not yet any clear or adequate explanation of when or how these familiar classical forms of life and language can emerge.

There is a similar problem in the cognitive neurosciences. There is a vast amount of detailed data on the physical brain states that modern cognitive science can show correspond to our subjective mental states. The problem here is that using *only* this objective physical detail, there is not yet any clear or adequate explanation of our personal subjective mental states that Hippocrates is talking about. At the root of both these problems is the necessity of separating the subject and object, or more precisely, separating our subjective and objective *models* of events. In fact this separation is necessary in order to give meaning to the concept of an *empirically testable model* that is required for any scientific model.

1 Epistemology—The Subject-Object and Symbol-Matter Relation

My first papers on laws and life present the view that the subject-object relation first arises with the emergence of *records* of events. Records require some form of material *symbols* that represent the events and an agent that interprets the symbols. This largely arbitrary symbol-matter relation first appears with evolvable self-replication, which I define as the origin of life. The emergence of symbolic function arises only when *useful* information is recorded by an agent. I explain why an adequate explanation of life depends on an epistemic process that has the essential function of a quantum measurement.

I also explain why evolution requires a *symbolic language* that physical laws do not construct and cannot interpret. This is the basis of the new field of biosemiotics. The genetic language appears to be the only language necessary for all of evolution, while human language, including mathematics, is the only language we have to express, communicate, and interpret the genetic language, evolution, and physical laws, including quantum theory. I discuss only some of the basic physical and structural requirements and limitations of these two powerful symbol systems that I call complete *languages*, and describe the fundamental similarities of genetic and human languages, as well as their enormous differences. Today the neurosciences and cognitive sciences are a rapidly growing focus of language research. Joanna Rączaszek-Leonardi discusses some of these current issues in language study and in cognitive science in her commentary [Afterword, p. 295].

My later papers extend these principles to classical models of reality that require hierarchical levels, with the levels being defined by complementary models. I also extend and elaborate the physicists' requirement that the universal *subject-object* and *symbol-matter distinctions* be separated by the empiricist's *epistemic cut* that appears conceptually inescapable in order to distinguish what we call objective knowledge from the individual subjective records or experience of the observer. This separation is an epistemic necessity, not a dualist ontology. I consider that the most primitive epistemic cut happened at the origin of life which separated the individual cell's genetic *informational* constraints from the objective lawful dynamics it controls. The concept of the *self* or the individual agent I define by the concept of *semiotic closure* that originates with the concept *self*-replication.

2 Where Did My Interest in the Physics of Life Begin?

The references in my papers give the published sources of many of my ideas, but there are deeper influences from my many teachers and colleagues. My memory is not reliable enough to recover all of these influences, but I clearly remember my most influential early teacher, Dr. Paul Luther Karl Gross, the Headmaster of my first boarding school. You will see why. He was a physicist born in Timbuktu where his parents were missionaries. His PhD dissertation was on the stackability of tetrakaidecahedra, (Gross 1927) a mathematical problem that also interested physicists like Lord Kelvin and biologists like d'Arcy Thomson. Dr. Gross taught us science, not as separate subjects like physics and biology, but as a way of thinking about what we observe. Instead of normal texts for our grade level he gave us original writings of great scientists and philosophers. Many of his lessons had a lasting effect on my thinking.

For example, Dr. Gross had us read the part of Plato's Meno where Socrates uses one of Meno's uneducated slave boys to support his idea of innate knowledge. Socrates draws a square in the sand and asks the boy to double the area of the square. After the boy sees it is more difficult than he first thought, Socrates inscribes another square of half the area forming four identical right triangles which the boy sees gives the solution. We caught on that Socrates' diagram was actually providing the solution; so Dr. Gross suggested that we try to demonstrate Pythagoras' Theorem by rearranging the triangles. With a few hints this allowed us to discover the elegant symbol-free direct perception proof, which we did not doubt was our own discovery. We also read Plato's allegory of the cave, and questioned whether forms like triangles were forms in nature or only shadows in the mind. By blowing soap bubbles in beakers we convinced ourselves that there was a physical reality to the irregular tetrakaidecahedra that formed as natural structures, not only in our bubbles but in the universe in foams, rocks and living tissue. Triangles, on the other hand, we did not find to be common natural structures. As I recall, we never reached an agreement on the ontology of triangles, but what I learned is that what we think is real is not a clear and distinct idea.

For summer reading Dr. Gross gave me the 1937 Everyman edition of Karl Pearson's *The Grammar of Science* (the first edition was published in 1892). Much of Pearson's discussion was beyond my comprehension, but Pearson's sensorimotor-based idealism and critical thinking about our mental constructs and limited modes of perception opened up for me an entirely new way of looking at the world. Like all children I began with a naïve realist outlook and never thought about how our senses, our brains, and our language affect what we tacitly accept as "out there" in the world. Years later I read the essay by the physicist Max Born (1969), *Symbols and Reality*, and I recalled that while reading Pearson's *Grammar* I had experienced the same shock that Born describes in his essay: "Thus it dawned upon me that fundamentally everything is subjective, everything without exception. That was a shock." Born went on to point out that: "Symbols are the carriers of communication between individuals and thus decisive for the possibility of objective knowledge." The physicist's concept of "objective knowledge" means only that knowledge that appears the same for all conceivable observers, as tested by the invariance and symmetries of the symbolic expressions of laws. Of course this concept of "objective" exists only in the physicist's brain and can be tested only by experiment.

What has motivated much of my thinking over many years is a question at the end of Pearson's *Grammar* in the chapter on Life, subtitled, "The Relation of Biology to Physics," Pearson's question is this: *"How, therefore, we must ask, is it possible for us to distinguish the living from the lifeless if we can describe both* *conceptually by the motion of inorganic corpuscles?"* The logical necessity of this question made sense to me. Everyone can see that life is strikingly different from inanimate matter; but clearly it is not an explanation of this difference simply to describe it by the same laws that describe inanimate matter. Pearson, after much speculation, admitted that he could not answer the question, and he concluded that the goal of understanding how physics relates to life "is at present indefinitely distant." Pearson's question is just a more specific way of asking, "What is life?"—but the form of his question makes it clearer why physicists are more troubled by the question than are biologists.

I wondered at the time why Dr. Gross gave me such a scientifically "outdated" book. When Pearson wrote the first edition of the *Grammar*, quantum theory and relativity were still a decade in the future, the ether was believed to be a conceptual necessity, and enzymes were still just mysterious "ferments." Many years later, I learned that *The Grammar of Science* was the first book Einstein read with his small group of friends called the Olympia Academy, a group founded by Einstein just before his 1905 "miracle year." Pearson's sensorimotor idealism was reflected in Einstein's thought. In his response to Jacques Hadamard's (1954) request to describe his creative thought process Einstein said, "The words or the language, as they are written or spoken, do not seem to play any role in my mechanism of thought." Instead, he said his concepts begin with "visual," "motor," and "muscular" elements, and only "laboriously" are they expressed in symbols so they can be communicated to others.

Dr. Gross also kept us up-to-date with current science. We sometimes attended Linus Pauling's evening lectures at the California Institute of Technology, which were popular science lectures open to the public. Without knowing anything about quantum mechanics, I learned that the periodic table, and in fact all of chemistry, could be explained by quantum mechanics. It was at one of these lectures that Pauling described Schrödinger's cat paradox. I could not understand that if quantum mechanics could explain all of chemistry and most of physics, why it also produced such obvious nonsense. It was a decade later in graduate school that I begin to understand why the creators of quantum theory felt the same way. This was the motivation for my early work (papers 1 and 4) in which I gave a reason why the classical results of measurements must begin long before cats. I explain why classical behavior must instead begin at the origin of life; but I do not address the formal part of the quantum measurement problem. Even today, the formal and conceptual incompleteness of quantum theory for explaining our classical experiences is one of the irritating unresolved foundational issues of physics.

3 The Symbolic and Subjective Aspects of Measurement—The Epistemic Cut

Observation or measurement is the ultimate test of a scientific theory. The general concept of measurement is a specific interaction of a measuring agent or instrument with a physical system that entails a symbolic outcome. As Born pointed out, it is symbols that allow communication between individuals, and that is decisive for objective models. By a symbol I include any *arbitrary* local structure that "stands for" or "refers to" the *specific* event or observable that is measured. The choice of what to measure and establishing the referent relation requires a subjective observer, agent, or interpreter. I did not know until graduate school that von Neumann (1955, pp. 418–421) had explained why laws, even in classical theory, could not usefully or *functionally* describe a measurement or an observation. One reason is that laws cannot determine what or when to measure or observe. That is a choice requiring a subject or agent. Another reason is that measurement is a specific selective act that must ignore most details. A fundamental condition for both quantum and classical physical theories is that the universal laws that apply to all details and all systems must be separate from the measured initial conditions that define a specific system.

Physicists, including Heisenberg, Schrödinger, Pauli, Von Neumann, and Wigner often referred to this conceptually necessary, but largely arbitrary, subject-object separation as the *Schnitt* or "cut" and in my later papers I call this unavoidable conceptual separation of the knower and the known, or the symbolic record of an event and the event itself, the *epistemic cut.* Schrödinger discussed this necessary conceptual separation of subject and object in many of his essays. Wigner (1982) felt that recognizing the necessity for this separation of laws and initial conditions was "Newton's greatest discovery." John Bell, whose theorems allowed tests of quantum non-locality, aptly called it the "shifty split." Bell (1990) also introduced the disparaging acronym, FAPP, meaning "for all practical purposes," which is itself a shifty epistemological concept that arises in the interpretation of all statistical theories including decoherence theory. Bell along with most physicists would not accept a theory if good "for all practical purposes" were its only test. It is important to understand that this separation of subject and object, however arbitrary and shifty, is a physical and logical requirement to make sense of any measurement, record, observation or empirically testable theory.

4 The Quantum Measurement Problem

What physicists call the *quantum measurement problem* has a long controversial history (e.g., Wheeler and Zurek 1983). The measurement process is considered one of the weakest areas of quantum theory. A quantum measurement is an interaction of a measuring device with a system described by quantum laws that the subject or observer sees as a classical outcome. There are three parts of the general measurement problem. The first part is the logical problem that, as von Neumann explained, also holds for classical theory. The problem is that the measuring device and the system it measures cannot be described by the same detailed model; otherwise the measurement becomes an indistinguishable part of a larger system, so there is no way to separate the result of the measurement from all the other interactions.

The second part of the measurement problem is what many physicists consider a fundamental *incompleteness* in the formalism of quantum theory. The problem is that quantum theory cannot *formally* describe a measuring device that must deliver a classical result. Again, it was von Neumann (1955, p.351) who made a clear statement of the problem. The measurement intervention *appears* to be instantaneous (non-dynamic), irreversible, and the physical representation of the result is largely arbitrary (e.g., pointer readings, symbols). At the same time, all microscopic events are assumed to obey reversible quantum dynamical laws (e.g., Schrödinger's equation). It was this inadequate model of measurement that prevented any formal decision about the state of Schrödinger's cat before it is observed.

The third part of the problem is that in neither classical nor quantum theory is there an unambiguous definition of the subject, agent, or observer that must determine what is measured and where to place the epistemic cut—hence Bell's phrease, the shifty split. By elaborating on von Neumann's (1966, pp.74–87) *logic* of self-replication, I define the *physical* conditions for a self-replicating unit that can function as a measuring agent. I discuss the epistemic cut in several papers, most thoroughly in paper 14, "Artificial Life Needs a Real Epistemology." Of course the "cut" is a metaphor that oversimplifies the essential complex relation between a model and the reality it represents, which is what epistemology is all about.

The Schrödinger cat thought experiment was a logical extension of the 1935 Einstein, Podolsky, and Rosen (EPR) thought experiment based on Einstein's concept of reality, demonstrating the "spookiness" of quantum coherence, a property Schrödinger described as entanglement (*Verschränkung*). Wigner carried the argument a step further by imagining a larger quantum system, the cat and "Wigner's friend" with a joint wave function, so that until Wigner hears a "yes" or "no" from his friend, the cat remains in a superposition of both dead and alive states. The early Copenhagen interpretation of quantum theory resulted in many physicists originally invoking consciousness of a human agent as the ultimate epistemic cut that terminates what otherwise would be an endless regress to larger and larger coherent quantum systems. My argument (papers 1, 4, 10) proposed that *functional* classical behavior results from a process equivalent to quantum measurement that begins with self-replication at the origin of life with the cell as the simplest agent.

None of my papers address the more difficult second part of the problem of relating formal quantum and classical models. It was not until the 1980s that decoherence theories developed suggesting how the appearance of classical behavior might be derivable from quantum theory by entanglement of the quantum system with the unavoidable "noisy" environment that FAPP, or more precisely, for all *measurable* purposes, obscures wave function interference (e.g., Zurek 1991; Schlosshauer 2006). However, even a successful decoherence theory would not adequately address the third problem of measurement because, in addition to decoherence, measurement requires an agent that separates *useful* records of events from the event itself by making an *epistemic cut.* Decoherence in various forms and degrees must occur everywhere in the universe at all times, but measurement only occurs when an agent *interprets* a specific sufficiently decoherent event as *useful* information. The problem is that *interpretation* and *usefulness* are subjective concepts. Decoherence has been interpreted as loss to the environment of unusable or undecipherable information. Decoherence has also been interpreted as a Darwinian selection process (Ollivier et al. 2004). I do not agree with this interpretation of quantum decoherence. It may be a form of selection, but it does not satisfy the evolutionary necessity of individual agents with heritable non-dynamic memory storage of open-ended capacity that allows the emergence of novelty and complexity.

5 History of Relating Physical Laws and Life

Throughout history the most common explanation used to distinguish living from lifeless matter has been some form of dualism, vitalism or divine creation. These are still the most common beliefs among non-scientists. Even in the middle of the twentieth Century many physicists seriously considered the possibility that life was so peculiar that it would require “biotonic laws” (Elsasser) or at least modifications of physical laws. Many of these views were as mysterious as vitalism, but at least they were considered to be within the reach of scientific study. Quantum theory appeared to open up a new realm of problems and possibilities. Parallels were drawn between the quantum measurement process and mind-body dualism by the founders of quantum mechanics, including Bohr, Schrödinger, Heisenberg, Pauli, and Wigner. It was the belief that human consciousness ultimately collapsed the wave function that produced the problem of Schrödinger’s cat and Wigner’s friend. It also supported the near solipsistic view that consciousness creates reality. From an evolutionary perspective this is difficult to imagine—because consciousness appears only as the latest emergent property of highly evolved life. Even today, J. A. Wheeler views reality as a “participatory universe” in which no phenomenon is a phenomenon until it is observed.

Physicists including Bohr, Schrödinger, Delbrück, and Wigner extensively discussed their doubts about the adequacy of quantum laws for explaining life. Bohr (1958) suggested that detailed experiments “will exclude the possibility of maintaining the organism alive.” Delbrück (1949) speculated: “It may turn out that certain features of the living cell, including perhaps even replication, stand in a mutually exclusive relationship to the strict application of quantum mechanics, and that a new conceptual language has to be developed to embrace this situation.” Wigner (1967) presented an argument that the probability of self-replication was vanishingly small assuming the linearity of quantum mechanics. I would say that Bohr, Delbrück and Wigner were correct that a strict application of *only* quantum mechanics cannot describe life, but not for the reasons they suggested. As I explained in papers 1 and 4, the reason is that replication requires a *measurement* process that quantum theory cannot describe.

The attitude of many physicists 70 years ago towards biology was characterized by Gunther Stent (1966) in the following words: “Thus it was the romantic idea that ‘other laws of physics’ (Schrodinger) might be discovered by studying the gene that really fascinated the physicists. This search for the physical paradox, this quixotic hope that genetics would prove incomprehensible within the framework of conventional physical knowledge, remained an important element in the psychological infrastructure of the creators of molecular biology” (p. 4).

6 The Rise of Molecular Biology—Recognizing the Symbol-Matter Problem

I do not remember when I began to understand why symbolic information could not be understood as only material structures obeying laws. When I started graduate school at Stanford University the genotype-phenotype distinction had been taken for granted by biologists for over 60 years and it was not considered an epistemological problem; although how the gene's symbolic information controlled material construction was not yet understood. Five years earlier, Beadle and Tatum had published the *one-gene one-enzyme* hypothesis. I knew nothing about it, even though they were working on the other side of the sandstone wall that separated my laboratory in the physics department from theirs in the biology department. There was no door between the departments, and at the time neither the biologists nor the physicists had any interest in constructing one.

That attitude changed during the following decade of the 1950s. In fact a revolutionary change took place with the rise of molecular biology. With Watson and Crick's discovery of the double helix the skeptical attitude of most physicists and biologists simply disappeared. Life became accepted as just ordinary physics and chemistry. No one was bothered by Pearson's question, and that attitude persists even today among most molecular biologists. The focus of physics began to shift to high energy particle physics.

That decade also saw the rise of what became Information Theory following Shannon and Weaver's 1949 publication that they properly called *The Mathematical Theory of Communication*. The general concept of 'information' was quickly adopted by physicists and biologists in many contexts. Much earlier, Maxwell's demon had shown that information could not be understood as just ordinary physics and chemistry; but even though Shannon made it clear that "communication theory" did not address the meaning or use of information, the common failure to distinguish the syntactic, semantic, and pragmatic aspects of symbols has resulted in much ambiguous and confusing literature.

I was first introduced to working molecular biologists at the Study Program in Biophysical Science at Boulder, Colorado in the summer of 1958 that brought together active physicists and biologists. Most of the discussions I remembered were about molecular structure. Because my research was in x-ray optics I was struck by John Kendrew's x-ray diffraction models of myoglobin. These meetings were probably where I first heard genetic language compared with human language. The neurophysiologist Francis O. Schmidt (1961, p. 28) saw, "a direct homology between the logic of molecular and phonemic codes."

The following year I spent at the Cell Physiology Department at Karolinska Institute in Stockholm with the intention of using x-ray microscopy to study cell structure. Instead, my interest shifted to the origin of life. The shift started during discussions with Joshua Lederberg on how to detect life on Mars, if it exists. The problem was to design instruments for NASA's first Mariner mission to Mars. I was also intrigued by Lederberg's suggestion in his Nobel lecture that it was the *information* in the DNA base sequences that determines the *folding* and therefore the function of the enzyme. This idea was confirmed by Christian Anfinsen's experiments on ribonuclease the following year for which he later won the Nobel Prize. I saw this as a primeval example of the symbol-matter relation—symbolic information controlling material function.

My first idea for a molecular model of replication came from the work of Giulio Natta on stereotactic copolymerization, and from the fact that many polymers grow as helices. I briefly describe this model in the paper 1 (1968). However, this model was too simple to have any evolutionary potential—a problem that is common to most origin of life theories. The original paper is archived online (see Pattee 1961).

The year before I was in Stockholm (1958), Beadle, Tatum, and Lederberg had won the Nobel Prize in Physiology and Medicine. It was only because Lederberg delayed delivering his Nobel lecture by 6 months (while he moved from the University of Wisconsin to Stanford University) that I was able to attend his lecture. That winter I also attended Arthur Kornberg's Nobel lecture when he won the 1959 Nobel Prize with Severo Ochoa in Physiology and Medicine. These two lectures were fine examples of the symbol-matter complementarity that Gunther Stent called the "informational" and "structural" approaches to molecular biology—the "one-dimensionists and the three-dimensionists." Lederberg as a geneticist emphasized the one-dimensional *information* in the DNA base sequences that determines function. In contrast, Kornberg as a biochemist focused on the three-dimensional *structures* of DNA and his polymerase enzyme. Of course it was clear that without enzymes the DNA is essentially inert, while without DNA there would be no enzymes; but what intrigued me was that a simple one-dimensional sequence in a molecular *structure* could function as *symbolic information* that controlled a very complicated lawful enzyme dynamics.

7 Why Life Requires Quantum Measurement

I was also intrigued by the converse problem of how enzyme dynamics could relate to information. I could not imagine how the quantum dynamics of enzyme catalysis could interact with specific products that function as classical information without involving what amounts to a quantum measurement. I assumed that the *dynamics* of enzyme reactions, like Kornberg's polymerase, ultimately require a quantum mechanical description, while the end result of this quantum dynamics must be Lederberg's classical, non-dynamic information. The enzyme's substrate-product relation also exhibits both the *arbitrariness* and the local *specificity* characteristic of the symbol-referent relation—what Monod (1971) called, "the principle of chemical gratuity." This arbitrariness implies the necessity of adaptors or codes that in physical terms are a special kind of non-integrable constraint. In paper 1(1968), I concluded that "the characteristic sign of biological activity at all levels is the existence of efficient and reliable codes." I imagined the cell as the agent that uses or interprets this type of information for survival. This suggested that the process of *self-replication* implies an *epistemic cut* long before consciousness. The early form of my argument is in papers 1(1968) and 4(1972).

I presented this argument that enzymes *in cells* perform the epistemic equivalent of quantum measurements at a small meeting on quantum theory at King's College, Cambridge, organized by David Bohm and Ted Bastin (Pattee 1971). As I recall, there was no strong rejection or acceptance of the idea during my talk, but at lunch Otto Frisch asked if I had resolved the reversibility of enzyme reactions with the irreversibility of the measurement process. Of course I had not, because the origin of reversibility is at the core of the measurement problem—quantum laws are reversible and measurement is irreversible. I had no idea of how to approach the problem beyond von Neumann's (1955) discussion. However, someone did point out that enzyme reversibility holds only in closed systems, which a replicating system is not. I later discussed my argument with Eugene Wigner and with Bohr's colleague, Léon Rosenfeld, who was a staunch supporter of Bohr's interpretation of quantum-classical complementarity, which some felt simply avoided the problem.

Wigner, who had long argued that consciousness was the ultimate cut, was apparently convinced by my argument. He wrote: "I believe I understand your arguments in this regard and concur with you. The reason for my arguing on the basis of consciousness was indeed that in this case I could adduce evidence for the incompleteness [of quantum theory], whereas I could not do this at a lower level" (note 1, paper 3). I'm not sure that Rosenfeld was ever fully convinced. After a year of correspondence he cautiously replied, "I did not meet any statement with which I would disagree," but he kept pointing out that my observation was not relevant to the measurement problem as he saw it. Of course my argument does not address the formal second part of the problem. Deciding *where* a classical description is functionally necessary says nothing about *how* it arises from quantum laws.

This view, however, provides one answer to Pearson's question. It shows that detailed quantum mechanical laws (such as those that describe the motion of his "corpuscles") cannot *alone* distinguish living from dead matter, because life depends on *classical symbolic records* and memory, and this requires measurement that is not derivable from quantum laws. It would also explain why the life of Schrödinger's cat does not have to wait for a conscious human observer, but is decided by the cat's internal measurements, as was the lives of its ancestor's going back to the origin of life.

Schrödinger in *What Is Life?* emphasized that genes must be individual, non-statistical, quantum mechanical molecules. This is true, but it does not address Pearson's question because *all* "corpuscles" living and lifeless are ultimately quantum mechanical molecules. Among several points that Schrödinger did not make clear is that *classical behavior is necessary for replication*. My view also suggests a plausible answer to what Von Neumann (1966) described as ". . . the most intriguing, exciting, and important question of why the molecules or aggregates which in nature really occur [in organisms] are the sorts of things they are, why they are essentially very large molecules . . ." (p. 77). I suggested that enzymes are small enough to take advantage of quantum coherence to attain the enormous catalytic power on which life depends, but large enough to attain high specificity and arbitrariness in producing effectively decoherent products that can *function* as classical structures.

There are still many ambiguities in this view, depending on the interpretations of 'decoherence.' For example, if one assumes that quantum theory is fundamental and that decoherent classical models are only good *approximations*, then one could say that fundamentally we are not alive; we are only alive FAPP—for all practical purposes—a statement that sounds at first almost as nonsensical as Schrödinger's cat. However, on second thought practicality makes more sense. The existence of life and the survival of individuals and species clearly depends on very practical observations and controls that are statistical. We do not experience wave functions, and quantum laws do not predict or guarantee life.

In the first paper (1968) you may detect my youthful impatience with molecular biologists in the 1960s who simply asserted that the physical basis of life is no longer a mystery because it obeys known physical laws. They clearly had no appreciation of Pearson's question. Of course, what constitutes a mystery is largely subjective. In paper 5 (1972), on hierarchy theory, I quote a famous molecular biologist who asked me, "Why do you need a theory of biology when you have all the facts?" This attitude is still common among many molecular biologists who are content with viewing life as just complicated but ordinary chemistry. This is an attitude toward mystery on which biologist and physicists profoundly differ. Biology requires models at many levels, and there is still plenty of mystery at all levels. Classical models are explanatory for most biological structures and for much of chemistry. Unfortunately, it is not yet computationally practical to treat large molecular systems entirely quantum mechanically. Consequently, chemistry simplifies molecular dynamics by treating only a few electronic degrees of freedom quantum mechanically and uses a classical approximation for all the rest. That is why most of chemistry is effectively a quasi-classical discipline separate from elementary physics. This leaves plenty of uncertainty about quantum dynamics, and contributes to the lack of consensus on how to decide when quantum mechanical description is useful or explanatory. Lacking convincing criteria and evidence for quantum effects has allowed room for speculative models, which have been discussed for years at many levels from cells to brains and by many authors, e.g., London, (1961), Fröhlich (1975), Conrad (1994), Hameroff and Penrose, (1996), Kauffman (2008).

Today, several decades after my first papers, there are many experiments suggesting that the rates of enzyme catalysis require quantum effects like tunneling (e.g., Allemann et al. 2009). There are also papers with proofs supporting my proposal that life would be impossible in a strictly quantum world (e.g., Prashant and Chakrabarty 2007). These proofs follow from the *no-cloning theorem* that quantum states cannot be replicated, which arose in the context of quantum computing (Wootters and Zurek 1982). The limit of quantum coherence is still an open question. Photosynthetic proteins apparently can control the rate of decoherence, and coherence has been shown to occur over surprisingly large distances, even with large molecules containing hundreds of atoms (e.g., Gerlich et al. 2011; Nairz et al. 2003). There are other plausible quantum coherence-dependent processes like kinetic and conformational proofreading. Even quantitative protein folding models may require quantum coherence at initial stages with decoherence developing at later stages. For a popular review see Vedral (2011). In any case, I think it is fair to say that at present no one should feel secure with any of the interpretations of decoherence or measurement in quantum theory. Pearson's question still needs much clearer answers.

8 Beyond Decoherence—The Autonomous Agent and Semiotic Closure

In none of my early papers did I make clear that the *sufficient* conditions for measurement are not addressed by the *necessary* conditions for measurement, including the decoherence process. That was partly because quantum decoherence was not a popular subject until the 1980s; but it was also because the third measurement problem of agency is a general epistemological problem requiring a subject and object distinction and the concept of *useful* information. These are not exclusive problems of quantum theory but occur at all functional hierarchical levels. Quantum theory just adds formal incompleteness and conceptual ambiguity.

The basic reason that *objective laws*, both classical and quantal, cannot derive or *explain* measurement events is because *measurement is a subjective process.* Measurements must be performed by a separate individual system—the autonomous agent that I define by *semiotic closure* (e.g., papers 10, 11, and 13). Autonomous agents could be artificial, like adaptive robots, as well as natural, like cells and humans. Laws can in principle *describe* only the material details of the measuring device and of the agent, but as I explain, following von Neumann, such a detailed description cannot *function* as a measurement.

An agent forms an epistemic cut between the system being measured and the measured result that is used or recorded. Laws cannot determine where or when this cut occurs, nor can laws specify which observable is measured or the result. Unlike decoherence, measurement is a functional concept that implies a physical result that carries *useful* information for an agent. Defining useful information is a problem. It has the same subjective difficulty as describing useful energy, work, or what we recognize as function or meaning. Definitions of what is useful energy and 'useful information' are context-dependent and differ in mechanical, electromagnetic, chemical, thermodynamic, computational, biological, evolutionary, social, economic, cognitive, and metaphysical contexts.

I avoid the difficulty of trying to define all these hierarchical levels of *function*, *utility,* and *useful information*, by focusing only on the most primitive cases, in the context of evolution, where *utility is first measured by its heritable survival value*. Obviously, survival is a precondition for all forms of utility. It is often assumed that natural selection ultimately determines survival value of genetic information at all levels of evolving organizations, but over cultural time scales this cannot be tested. Although social and neural Darwinism (e.g., Edelman 2006) may explain some aspects of behavior and thought, species survival is not the selective criterion at higher level cognitive models in brains. In human brains there are cultural, rational, aesthetic, and other artificial selection criteria that appear to evade or sometimes override genetic natural selection. For example, scientific criteria for good models do not depend on natural selection but on cultural selection—including how well the models conform to observations, as well as other logical and aesthetic criteria.

It is one of the central points in my papers that the earliest place that an epistemic cut can be objectively defined is at the first level of organization where information is *useful* or functional *for an agent*. The concept of model also implies that an agent must construct and use the model. I discuss the modeling relation in detail in paper 15, "The Problem of Observables in Models of Biological Organizations." In practice most sciences proceed without worrying about epistemology and where to place the epistemic cut because the human scientist is the agent who decides all this, often tacitly.

What are the conditions for an autonomous agent? It follows from von Neumann's *logic* of self-replication that both the construction process and the description-copying process need to be coded in the stored informational description. Von Neumann's logical theory of self-replication did not address the physical requirements for implementing his logic, however. That is the problem that I call the *physics of symbols*. In order to *physically* implement this logical closure it is also clear that the symbolic instructions must have a material structure that constrains all the lawful dynamical processes of construction. I first called this self-referent symbol-matter completeness *semantic closure*; but at Luis Rocha's suggestion I now call it *semiotic closure* because this is the simplest level exhibiting all three aspects of the semiotic processes, syntax, semantics, and pragmatics—the genetic code and its translation are the *syntax*, the protein folding where the energy-degeneracy of the information is removed and a specific local rate control results is the *semantics*, and the consequent effects on the organism are the *pragmatics* of the stored information. The point is that *only if there is semiotic closure can an organized unit function as an autonomous agent.*

Von Neumann did see the necessity of a non-dynamic "quiescent description" distinct from the physical dynamics of construction. I describe his logic in the last paper, "The Necessity of Biosemiotics." What I call the simplest autonomous agent exhibits von Neumann's "threshold of complication" where open-ended evolution begins. It also corresponds to what Gell-Mann (1994, p. 155) and Hartle have called an IGUS, an Information Gathering and Utilizing System. *The self, whether a cell or a human, is a unique individual only by virtue of the classical memory patterns in its genes and in its brain.* In fact, the subjective *memories* in genes and brains are what define the self, the individual, or the autonomous agent. *Matter can function as a symbol only by virtue of an agent acquiring and using information to constrain a lawful dynamics.*

9 All Objective Models Are Also Subjective

As Hippocrates, Born, and many philosophers have stated so clearly, all our models exist only in individual brains. What physicists call an objective model is just a very restricted type of subjective model that is common to all subjects by virtue of invariance and symmetry relations *inherent in the symbols* of the model. I have extended this concept of subjectivity to the origin of life. The relation of an agent's internal subjective models to its external environment is the fundamental problem at all evolutionary levels. I have labeled these subject-object relations that require an epistemic cut between the agent's model and what it models the general *symbol-matter* problem. I argue that this problem exists at all levels from the origin of life to human thought. In artificial intelligence it is often called the symbol grounding problem; at the cognitive level it is the problem of reference. At the highest levels it is called the mind-body problem, which involves even more obscure problems like intentionality, qualia, and consciousness. The autonomy that requires semiotic closure is discussed in several papers including paper 10, “Cell Psychology: An Evolutionary Approach to the Symbol-Matter Problem.” In papers 13 and 15 I quote Heinrich Hertz (1956) who also understood that all models are subjective. He stated concisely the empirical criterion for the modeling relation that is accepted by most modern physicists.

> We form for ourselves images or symbols of external objects; and the form which we give them is such that the logically necessary (denknotwendigen) consequents of the images in thought are always the images of the necessary natural (naturnotwendigen) consequents of the thing pictured…. For our purpose it is not necessary that they [mental images] should be in conformity with the things [external objects] in any other respect whatever. As a matter of fact, we do not know, nor have we any means of knowing, whether our conception of things are in conformity with them in any other than this one fundamental respect.

The problem of how closely our images and symbols “*should be in conformity with the things*” is the crucial subjective and empirical issue at all levels. An engineering model may conform for all practical purposes (FAPP), but that is not good enough for most scientists. There the ideal of truth is expressed by Wigner’s feeling that “it could not be otherwise,” as well as the aesthetic conditions of simplicity and elegance to which a satisfying model must conform. The converse problem of how matter becomes symbolic in the beginning—the origin of life problem—is still an obscure issue. Once life begins, evolution produces higher levels of agents with more complex epistemic cuts and models, resulting in endless and unpredictable measures of fitness.

10 The Concept of Genetic Language

The only language I was thinking of in paper 1 was the “language of physics” meaning its technical vocabulary. At the time, many people were intrigued by the similar structures of human and genetic language, and as I recall I simply accepted the “genetic language” concept as it was used by geneticists like George Beadle and Joshua Lederberg, and by the linguist Roman Jakobson who called the similar structures of genetic and human languages “isomorphisms.”

It was the novel concept that molecules obeying laws could somehow be coordinated to form a language that first provoked my thinking about the physics of symbols. It was after I recognized the necessity of the description-construction duality required by von Neumann’s logic of self-replication that the “lone symbol” problem arose. It became clear that a useful description requires more than an isolated symbol. In the second paper (1969), “How Does a Molecule Become a Message,” I point out that symbols do not exist in isolation but only in coordinated groups. If symbols are to be rich enough for unlimited evolution as well as for creative thought, the symbols must belong to a *complete* coherent symbol system—which I call a *language*. I still do not understand the sufficient conditions for complete, open-ended, or creative languages, although I can discuss some necessary conditions. The problem is suggested by questions like: What function has a sequence of nucleobases until it constructs an active enzyme? What function has an enzyme outside a cell? What function has a cell in an organism? What meaning has a string of words until it forms a sentence? What meaning has a sentence without an interpreting agent? Of what use is the information to the agent without a society of agents? In other words, function and meaning are open-ended, context-dependent concepts. Joanna Rączaszek-Leonardi, in her commentary, discusses how language cannot be understood except in the context of an organism's dynamic interaction with its community and ecosystem.

The linguists' limiting definition of language as just what humans speak I felt was a circular unproductive definition. Without agreeing with their theories of language, I was influenced by the concept of the universals of language as studied by Zellig Harris (1991), Charles Hockett (1960), and Joseph Greenberg (1978). Also influential were opinions of the biologist François Jacob, the linguist Roman Jakobson, the neuroscientist, Francis. O. Schmitt, and the geneticist Hans Kalmus. Martin Sereno (1991) has discussed specific analogies between genetic and cultural languages.

The genetic language has the expressive power to construct novel organisms and to control novel behavior of unmatched variety. Human language has expressive power to communicate, influence behavior, and endlessly create novel thoughts. These two languages are the only ones known to have such open-ended expressive power. Consequently, whatever is common in these two languages, I consider to be of fundamental significance. The differences are great and also significant as one would expect after 4 billion years of evolution.

11 The Necessity of Complementary Models

In all of my papers I regard complementary models as an epistemic necessity. The inescapable fundamental complementarity is between the subjective and objective models of experience. This is a universal and irreducible complementarity. Neither model can derive the other or be reduced to the other. By the same logic that a detailed objective model of a measuring device cannot produce a subject's measurement, so a detailed objective model of a material brain cannot produce a subject's thought. Physical laws are objective by virtue of symmetry and invariance principles that guarantee their independence from the subject's physical state or world view. At the same time, every subjective agent is identified as an individual by its unique local view. In fact, the concept of invariance with respect to individual states would have no meaning without the existence of different individual states.

Complementary models as I define them are models of a system that may be *formally* incompatible in the sense that no one model is logically or mathematically derivable from, or reducible to, the others, and all such models are necessary for a complete understanding of the system. Of course, what constitutes a "complete understanding" is largely a subjective issue. It depends on how you choose to isolate and define the system, exactly what aspects of the system (what observables) you choose to model, and how precisely you choose to model them. I emphasize that the formal incompatibility of two models is only an epistemic necessity. *It does not imply dualism.* It says nothing about ontology and does not exclude conceptual compatibility—although this is often difficult because conceptual categories such as 'discrete' and 'continuous' derive from different pattern recognizing regions of the brain. For example, even though the reversible laws of particle dynamics are *formally* incompatible with irreversible laws of statistical mechanics, we can picture the transition conceptually from one model to the other by imagining the number of particles gradually increasing and our detailed information gradually decreasing. Evolution has produced brains with the remarkable ability to coordinate many types of independent sensory patterns such as sight and sound into a coherent mental image.

There are also other reasons for the necessity of complementary models. First, in physics there are specific existing cases—e.g., the classical complementarity of formal reversible and irreversible equations, and quantum theory's *intrinsic* complementarity of discrete particle and continuous wave aspects of matter and energy. Second, in biology the *structure-function* complementarity is a universal necessity. Function cannot be logically derived from only a structural description, and a structure cannot logically be derived from only a defined function. In evolutionary terms, structure-function relations appear to be "discovered" by natural selection and often appear as frozen accidents. Third, animals recognize discrete objects and continuous motion in separate regions of the brain, and neither region would make sense without the other. For humans the description of reality by categorical dichotomies (e.g., alive versus dead) is enhanced by linguistic conventions—but these distinctions must have arisen from how the brain recognizes patterns. Fourth, structure-function complementarity occurs in language in the relation between syntax and semantics. One cannot derive meaning from syntax, nor can one derive syntax from meaning. Fifth, complex systems require hierarchical models. In physics a change of scale in space, time, or energy of several orders of magnitude usually requires new observables and a new model. Existing physical models cover space scales from the Planck length, ~10^{-35} m, to the observable size of the universe, ~10^{26} m. In papers 5 and 6 on hierarchy theory I explain that different levels of a single system are distinguished as such because they require different models. "Complementarity" appears in the title of only two papers, paper 8 (1978), "The Complementarity Principle in Biological and Social Structures," where its history and meanings are discussed, and paper 17 (2007), "The Necessity of Biosemiotics: Matter-Symbol Complementarity," but the concept is essential in all of my papers.

In spite of this evidence, the idea of the necessity two irreducible and often inconsistent models is often rejected on philosophical grounds. It still surprises me that the concept is so controversial, because it is evident that one universal model of reality does not exist, and everyone uses more than one model of their experiences even in everyday life. Nevertheless, much of the literature in many scientific fields is full of unproductive arguments over which of several complementary models is correct or superior to the others. Complementary models are also rejected by the philosophy of reductionism that is still common in molecular biology and much of the neurosciences. Another objection to complementarity is the logician's abhorrence of contradiction, but such contradiction, properly understood, can occur only within a single formal model.

The term complementarity was made popular by Niels Bohr. Unfortunately Bohr's description of complementarity was obscure, as were many of his statements, but he was trying to explain the quantum wave-particle complementarity, and that is inherently obscure. The general concept goes back at least to Heraclitus' upward-downward path and to Aristotle's four complementary causes. Nicholas of Cusa was more explicit with his *coincidentia oppositorum*—unity of opposites. Euler pointed out that the Creator's laws can be described equally well by time-dependent efficient cause (time-dependent state-determined equations) or final cause (timeless extremum principles). Hermann Weyl (1949) makes a convincing case for the necessity of subject-object complementarity in his *Philosophy of Mathematics and Natural Science*. He calls immediate subjective experience 'absolute' in the sense that we cannot alter it, and objective knowledge he calls 'relative' in the sense that individual experiences necessarily differ. Weyl says, "It seems to me that this pair of opposites, subjective-absolute and objective-relative contains one of the most fundamental epistemological insights which can be gleaned from science" (p. 116).

Bohr believed that complementarity "bears a deep-going analogy to the general difficulty in the formation of human ideas, inherent in the distinction between subject and object." But I agree more with Weyl; complementarity is more than an "analogy" to the subject-object relation in human thought. I would say that *the subject-object distinction separated by an epistemic cut is at the core of the general symbol-matter problem at all evolved levels beginning with the origin of life.*

Natural language has many complementarities. It incorporates many logically dichotomous concepts such as 'discrete' and 'continuous' that do not unambiguously conform to any elementary forms of nature. The classical example is Zeno's discrete simulation of continuous motion that failed to usefully describe motion. Making a conceptual transition from a discrete model to a continuous model is difficult because the brain has evolved different neural structures or codes for recognizing edges of objects and motion of objects. Aristotle recognized that motion was not conceptually discrete and concluded, "That which moves does not move by counting." It took nearly 2000 years for mathematicians like Cauchy, Dedekind, and Weierstrass to formally define continuity in terms of discrete symbols, although their definitions still require the limit of an endless process. Continuity is difficult to symbolize because it implies a concept of infinity. Cantor proved that a countable infinity implies the existence of an uncountable infinity, thereby confirming Aristotle's observation.

The complementarity of discrete and continuous models is a fundamental aspect of the symbol-matter problem. Evolution prepared the simplest brains to distinguish discrete objects from the continuous motion of objects, thereby allowing effective sensorimotor control. Our everyday experience as well as classical physics is based on a clear and objective distinction between discrete particles and continuous motion. In modern physics, however, this clear distinction is no longer possible. Discrete particles and continuous fields, matter and energy, space and time are no longer objectively separable but depend on how we observe nature. It appears that our artificial instruments have extended our senses beyond what our classical brains can model without cognitive dissonance. It is not clear how far we can reduce this dissonance by learning new concepts.

12 The Origin of My Interest in Symbolic and Dynamic Models

Another teacher who influenced my thinking about models was the mathematician George Pólya. My first serious efforts to analyze my own mental processes in solving problems began with mathematics classes taught by Pólya. Also influential were the arguments I had with Allen Newell, my closest friend during our first years of physics graduate study. Newell later became one of the founders of Artificial Intelligence with Herbert Simon. During our first years of physics graduate study (1949), physics courses were full of standard mathematical solutions with very little interpretation. The EPR experiment and Schrödinger's cat were not considered essential topics. Bell's theorem and non-locality experiments were 15 years in the future. All that was expected in physics classes was to get on with the math. The mathematics was difficult, so I did not think much about foundational issues. The only epistemic assumption was Bohr's condition that the entire experimental situation be expressible in "ordinary language." However, nobody considered what the requirements were for "ordinary language," or what role language played in thinking about physical models.

This approach to physics was a contrast to what Newell and I were learning from Pólya, whose interests beyond mathematics were the psychology of problem solving and pedagogy. Two of Pólya's heuristics were to "look at extremes" and "change representations"—by which he meant to look at a problem in as many alternative ways imaginable, using imagery, drawings, graphs, arrays, analogy, natural language metaphors—in fact anything other than the representation in which the problem was presented. He also stressed the value of intuition and guesswork. Pólya typically saw any solution to a problem as just one case of more general and more fundamental problems. "Variation of the problem is essential," he said. For example, in one class he generalized the Pythagorean Theorem to include not only squares on the sides of the triangle, but areas of any similar shapes whatsoever, and then to 3, 4, and n dimensions, and to infinite dimensional Hilbert space, the space of quantum states. At the same time, he placed great importance on metaphor, the choice of words and even the choice of individual symbols that while apparently manipulated entirely syntactically always convey a tacit semantics—a cognitive process that is totally absent within computers.

Pólya had us read Hadamard's (1954) *Psychology of Invention in the Mathematical Field*, which is a collection of reports from well-known mathematicians, scientists, and philosophers on how they think. Except for one case (that of the philologist

Max Müller) none of them begin thinking with ordinary language or even formal mathematical language. In fact many mathematicians and some philosophers regard natural language as misleading; like Berkeley who found words a great "impediment to thought," or worse, Schopenhauer who felt that, "Thoughts die the moment they are embodied by words." (quotes from Hadamard). It is generally agreed that creative activity begins at the ineffable level of preconscious thought that gradually, or sometimes suddenly, is brought into a visualizable or symbolic form. The question that has interested philosophers and now motivates neuroscientists and cognitive scientists is what is going on in the brain at this preconscious level and how do instincts and images lead to language and the formal syntactic level of mathematics.

Newell and I were thus confronted with these two points of view: In physics we were taught to compute and not think too much about interpretation—what David Mermin later characterized as the "Shut up and calculate" school. On the other hand, Pólya was stressing imagination and auspicious guesswork, leaving formal symbol manipulation as the last step. As he wrote: "When you have satisfied yourself that the theorem is true, then you start proving it." I believe we saw this as a 'problem' only because of the way physics was taught. Working mathematicians and physicists usually do a lot of vague thinking before they start writing.

The source of our lengthy arguments was that Newell and I did not think the same way, at least at the level where we could access our thought processes. To oversimplify, Newell felt his primary thinking was symbol manipulation, while I felt that mine was playing with images. But we both saw the common problem of how brains could exhibit both modes of thought as well as all the heuristic variety that Pólya and Hadamard discussed. The mystery was (and still is): How do neurons perform all these very different kinds of thinking? Our conclusion was that such different cognitive processes must occur at different levels of the brain's activity. This was the beginning of our thinking about the nature of hierarchic levels and about the complementary models necessary for each level. Newell later depended on hierarchical architectures in his programming languages, but he believed that the root level in the brain—as in computers—would be simple discrete symbols. He maintained this view throughout his life.

Indeed, there was a strong basis for his belief at the time. We had not yet seen a material computer, but we knew about the logical Turing machine and the remarkable *formal* concept of the *universal computer*. Neuron firing was also viewed as a discrete process, and the most popular model of the brain at the time was presented in the paper by Warren McCulloch and Walter Pitts (1943), "A logical calculus of the ideas immanent in nervous activity" in which they assume that, "The activity of the neuron is an all-or-none process." It seemed plausible that the brain at the neuronal level operated something like a "universal" logical network, and as computers developed, this became the most popular basis of artificial intelligence models.

As a physicist, I could not understand how classical symbols could be a primitive concept without an underlying quantum dynamics. I argued that discrete symbol vehicles must be physical structures that evolved to act as constraints on the fundamental continuous dynamical laws. For the entire year, while Newell and I were working on x-ray reflection optics as possible PhD research projects, we continued arguing about discrete and continuous models, and about what properties a physical structure must have to be recognized as a symbol (paper 7). This is still an unsolved issue in contemporary artificial intelligence models and brain models.

The next year Newell decided he did not want to do lab work and went to Princeton to study mathematics. We continued our discussion for over a year until Newell gave up mathematics except for the relatively new field of Game Theory—which he saw could be applied to human strategies. Newell finally got his PhD with Herbert Simon with whom he collaborated for the rest of his life. Simon was awarded the Nobel Prize in economics in1978, but he is probably best known for his work in artificial intelligence and in computational theory of human cognition. His paper, *The Architecture of Complexity: Hierarchic Systems* (Simon 1962) influenced my ideas on hierarchy theory, as I discuss in papers 5 and 6. Both Simon and Newell acknowledged Pólya's influence on their work. Many of their strategies like *protocol analysis* were based on Pólya's pedagogy and heuristics. In much the same way that Turing used the ways in which humans calculate to design his abstract machine, Newell and Simon used the ways in which humans solved problems to design their computer programs.

After developing many famous problem-solving programs and computer languages, Newell and Simon formulated the "Physical symbol system hypothesis" (e.g., Newell 1980) that states, "A physical symbol system has the necessary and sufficient means for general intelligent action." This is often equated with "the computational theory of mind." They were convinced that this hypothesis solved the mind-body problem. This approach persists today among many neuroscientists using a wider view of computation, one that now includes artificial neural nets.

I disagreed for two reasons. First, the hypothesis does not connect the dynamics of physical laws with non-dynamic symbol vehicles. Second, from my evolutionary point of view this hypothesis omits over a billion years of complex pattern recognition and dynamic actions in the simplest cells, and hundreds of millions of years of sensorimotor controls requiring direct perception of patterns, all of which must be accomplished in milliseconds. A flying insect does not have time to compute its flight pattern by programmed symbol strings. I regarded their hypothesis as only an ambiguous *instance* of the symbol-matter problem. It was not a solution of the problem.

Moreover, their hypothesis provides no basis for deciding when or how a material structure can be recognized as a symbol. It also says nothing about how symbols control matter or acquire meaning—what Stevan Harnad (1990) later labeled as the "symbol grounding" problem. Of course Newell and Simon were thinking of programmed computers as their physical symbol systems, and they demonstrated that computers could simulate some human behavior that we call intelligent. But the claims of "general intelligent action" turned out to be greatly overstated, and the programmed computer model was too far removed from how brains work. For many AI workers at the time, the idea of Turing *formal* equivalence was extended to the idea of *functional* equivalence, and this was often used to justify computer simulations that ignored the neural anatomy of brains and the physics of material symbol vehicles. I take up this topic in Paper 7.

13 The Evolutionary Limitations of Programmed Computer Models

A basic problem with computational universality is that it is a *formal* concept that does not capture the essential properties of physical laws, evolution, and language. The basic conditions required for a formal symbol system are that the symbols form a fixed set that can be read and written by strictly deterministic rules with absolute noise-free precision. Many proofs also require the concept of infinity. Formal Turing-universal equivalence exists for many symbol systems including artificial neural nets and cellular automata (e.g., Wolfram's Turing-equivalent CA Rule 110). However, formal equivalence assumes no noise or contact with an unknown environment. In cells and brains this could never be possible.

Evolution is a case where such conditions also can not be met, for nothing would evolve if they were met. In fact, evolution by natural selection depends on a noisy environment. The response to noise in the genetic symbols is what creates new functional units. 'New construction' does not exist in formal systems, and the consequence of noise in formal symbol systems is usually a meaningless result. Computational models of cognition exhibit the same lack of creative behavior. I discuss this problem in paper 13 (1995), "Evolving Self-reference: Matter, Symbols, and Semantic Closure."

There is also the related problem of the explicitness of computation, as contrasted with the vagueness, ambiguity, and guesswork typical of creative thought. At Waddington's second Bellagio meeting, the philosopher Marjorie Grene introduced me to Michael Polanyi's insights on the failure of all our symbolic expressions, even formal mathematical expressions, to achieve the ideal of objectivity. Polanyi's anti-reductionist arguments show how all of our explicit symbolic descriptions must be grounded in a reservoir of ineffable structures and subsidiary knowledge. What I call symbolic constraints on a lawful dynamics Polanyi (1968) calls special boundary conditions that "harness" dynamical laws. My hierarchy theory was an elaboration of Polanyi's concept of the functionally irreducible hierarchical levels of boundary conditions.

It was not until the 1980s that the more realistic (and older) concept of *concurrent distributed networks* was reestablished, along with their control of material robots that acted on continuous dynamic physical information embedded in real environments. It is now generally accepted that the brain's neurons do not process bit strings like a programmed computer; but how brains can calculate like a computer is still not understood. In paper 7, "Discrete and continuous processes in computers and brains," I show why both discrete and continuous models are conceptual necessities for complete description at all hierarchical levels. In paper 10, "Cell Psychology: An Evolutionary Approach to the Symbol-Matter Problem," I explain why the symbolic computational view and the non-symbolic dynamics of the Gibsonian "direct perception" ecological psychologists are complementary models. Pearson would call all of these views different modes of summarizing and discriminating groups of sense impressions.

14 Misconceptions of Physical Laws and Evolution Theory

In all my papers, I accept that physics and evolution have well-established, empirically tested models to which I conform in my discussions. For example, I assume symbols exist as material structures (symbol vehicles) that, as material structures, must obey physical laws. This implies that symbol systems, including genes, natural language, mathematics (as well as all forms of computation), must be compatible with classical and quantum mechanics, statistical thermodynamics, and information theory. Yet, this appears to conflict with the striking feature of all these languages—i.e., that in spite of obeying these inexorable laws, language appears to be completely undetermined by these laws.

This apparent conflict usually results from a common misconception of physics which holds that because laws are inexorable—that is, no event can disobey laws—the implication is that *laws determine all events*. *That is not the case.* The most fundamental complementarity in physics is between laws and initial conditions. Laws cannot derive initial conditions, nor can initial conditions derive laws—but both are necessary for experiments. Most of the structures in the universe are undetermined by laws or are accidental. Measuring instruments obey physical laws but are not determined by laws. Eddington (1929, p. 260) emphasized this fact in *The Nature of the Physical World*: "There is nothing to prevent the assemblage of atoms constituting a brain from being of itself a thinking object [including free will and consciousness] in virtue of that nature which physics leaves undetermined and undeterminable." Gell-Mann (1994, p. 134) again pointed this out in *The Quark and the Jaguar*: "the effective complexity [of the universe] receives only a small contribution from the fundamental laws. The rest comes from the numerous regularities resulting from 'frozen accidents.'"

Along with this misconception of physical theory, there are corresponding misconceptions of evolution theory. The fact that genes determine the amino acid sequences of all proteins does not imply that genetic information controls all biological events. Likewise, the fact that all organisms are subject to natural selection does not imply that natural selection is responsible for all of the structures of organisms. The fact that genetic information is a unique source of order that is crucial for evolution does not mean that it is the only source of order. Physical laws and many nongenetic, self-organizing, and developmental constraints are necessary at every stage of the execution of genetic information (e.g., Kauffman 1993). In paper 14, "Artificial Life Needs a Real Epistemology," I point out that computer models of life must clearly define the epistemic cut—i.e., specify what data represent the physical environment and what data represent the organism's *model* of the environment.

The best analogy I know to genetic language is the information found in the score of a symphony. The symbols of the score are essential for the music, but they are a very small part of the execution in an actual performance. The symbols of the score uniquely define the opus but the score indicates nothing of the complexity of execution and the effect on the audience. Every note requires complex dynamics, both in the physics of the instruments and the sensorimotor skills of the performers.

The audience, in turn, responds with unpredictable personal emotions. All these dynamics are symbolically ineffable. The same incompleteness is the case for genetic instructions. The analogy can be carried even to the creation of the score, which must have involved a process of search and a kind of 'aesthetic natural selection' in the mind of the composer. However, this analogy is still too simple and breaks down because of the complexity of epigenetic and developmental feedback networks that continually and strongly affect the expression of the genes. A conductor demands that all musicians are reading the same page of the score. An organism has no conductor, and each cell can turn to a different page of its genome depending on signals from neighbors and its environment.

15 Significant Differences Between Genetic and Brain-Based Languages

In several papers I have described the similarities that I believe are important between genetic and human language (e.g., Paper 11, "Universal Principles of Measurement and Language Function in Evolving Systems"). Some of the many differences are also important. The above analogy comparing genetic expression to the performance of a musical score does not work for the brain or with human language. Also, life must have begun at the molecular level with naturally occurring molecules that could grow as discrete strings. The written symbol strings of natural language are only the latest discovery of humans. Genetic language began with discrete strings of a small number of molecular symbol vehicles in single cells and ended up controlling the complex dynamical metabolic network of an entire multicellular organism. Conversely, human language began with a complex dynamical neural network and ended up producing discrete strings of a small number written symbols.

Little is known in detail about how memory is acquired, stored, organized, or accessed in brains, but there is no evidence of anything like a musical score or discrete symbols vehicles. Non-dynamic symbols appear only after motor outputs, as in writing. We know that recognizing even a single written symbol requires a coordinated dynamic pattern of millions of neurons. The primeval genetic symbol-matter problem is how non-dynamic discrete molecules become symbol vehicles that control chemical dynamics. We can follow the reading and control processes in the cell at the chemical level in detail. The symbol-matter problem in brains is the converse of the problem in genes: How does a complex dynamic neural network produce the concept of a discrete symbol that can end up written as simple arbitrary mark?

One important functional difference between these languages is that variation in *genetic information must be expressed before selection can begin*. Natural selection operates only through the phenotype, not directly on the genotype. Animal brains, in contrast, have the enormous advantage of being able to acquire, evaluate, and select information before expression. In other words, using models, animals can *predict before they act.* That is why brains evolved. It is at the human thoughtful planning level that the concepts of *choice* and *purpose* acquired their conventional literal meaning. That is why applying the concept of purpose to cells, even as a metaphor, produces more polemics than enlightenment.

16 Will Models of the Brain Improve the Brain's Models of the Universe?

Von Neumann in *The Computer and the Brain* (1955) raised a provocative question of whether more knowledge about the brain might alter the way we look at mathematics and logic. The wider question is: Should we expect more detailed models of the brain to help us interpret the brain's own models of the universe, and perhaps improve the models themselves, or even create more effective conceptual forms? For example, is it conceivable that understanding the brain might help us comprehend quantum behavior, or perhaps more likely, help us see why we cannot comprehend quantum behavior?

Von Neumann's question about mathematics has been indirectly addressed by others, such as Hadamard (1954), Pólya (1968), Wigner (1960), and McCulloch (1988), and more recently it has become the focus of studies by cognitive scientists (e.g., Lakoff and Núñez 2000; Dehaene and Brannon 2011). The neurosciences and cognitive sciences are the latest growth disciplines that are uncovering fascinating details of how the brain works. Few neuroscientists today think of the brain as processing symbols like a programmable computer. It is now generally accepted that even the most abstract thought has evolutionary roots in the body's sensorimotor interactions with the environment. But in spite of this *embodied cognition*, it is far from clear how much of the brain's detailed structure corresponds to useful models of the universe, and how much structure is arbitrary or incidental. Even for genetic sequences this distinction is still a difficult empirical and conceptual problem.

A few hundred million years of evolution has probably selected brains that work near optimum efficiency for basic survival function like sensorimotor control and pattern recognition. For the individual, knowing more details about the brain is not likely to improve these functions. External or objective knowledge of neural details may help alleviate some brain pathologies by the use of drugs or surgery. Brain-computer interfaces and neuroprosthetics now allow a practical form of psychokinesis. Neuroscientists are using exceedingly sophisticated brain imaging technologies to correlate regions of brain activity with sensory inputs, subjective thoughts and emotional states. New types of multichannel implanted electrodes can capture single neuron activity, and genetic and optical techniques allow control of individual neuron activity. Many cognitive scientists expect that detailed knowledge of how we learn will help improve pedagogy or how we teach.

However, no matter how much detail an external *objective* observer discovers about brain's activity, the *subject* must experience its events in a different way. The fact is that the objective details of thinking are not available to the individual doing the thinking, and the reason is that knowledge of these details would not be functional.

This is obvious with language. We could never produce a sentence if we were burdened with following a detailed model of the synaptic activity necessary to produce a meaningful sequence of symbols.

I believe this is a general structure-function complementarity principle. It is the reason that our models are hierarchical. Von Neumann showed why *ignoring* details is *necessary* for the *function* of measurement. Computer codes provide another clear example of the necessity of loss of detail at each hierarchical level in order to obtain increasingly efficient function. We all use computers without knowing what is going on in any detail, and knowing the structural details would only obscure the *function*. At the same time, we know that computer design and construction requires many *structural* levels of increasing detail, starting with solid state physics and quantum theory.

This selective loss of detail appears to be essential for any function, and the principle must also apply to brain models. For the subject, too much detail at any level will obscure the *subjective* brain's function at that level. However, a complementary, detailed structural model of neural activity will be necessary to understand *objectively*, e.g., how speech is generated. A single model cannot adequately explain both structure and function. As is the case with the simplest measurement in cells, a single model of the brain will not adequately explain both sides of an epistemic cut.

It is precisely because structure and function, as well as objective and subjective models, are logically and conceptually disjoint—i.e., neither type of model can be derived from or reduced to the other—that we cannot predict how knowledge of one will affect the other. How much we can expect objective knowledge about the brain to improve the brain's creativity is as unpredictable as is the course of evolution. All our thinking is affected by how sensorimotor control evolved, but that is not what is most important about large brains. The power of the human brain is in its creativity and artifacts. Humans long ago invented tools that extend the power of the human muscles by machines, speech that allows communication, writing that extends the power of memory, and many instruments that extend the power of the senses. Our unassisted brains even today are not powerful enough to have created even the most basic scientific constructs such as molecules, spacetime, wavefunctions, quarks and black holes without first having extended the senses by inventing sophisticated instruments and by learning the language of abstract mathematics. In particle physics, cosmology, genetics, and the neurosciences, computers have also become a necessary prosthetic device for developing models.

Unfortunately, in the evolutionary time scale, whether or not these cognitive models and their resulting Promethean technologies, like nuclear and genetic engineering, will ultimately favor the survival of the species is not at all obvious. In any case, as long as humans survive, these epistemological conditions will not disappear. Our models of living organisms will never eliminate the distinction between the self and the universe, because life began with this separation and evolution requires it. Neither can our models of the physical universe eliminate this distinction. That means we will still have to recognize epistemic cuts that separate subject and object, and construct complementary models for understanding laws, language, and life. We will never stop making models, and this means that neuroscientists and philosophers will face the never-ending task of elaborating Hippocrates' list of "what men ought to know" about the brain—where all our models reside.

References

Allemann, R. K., Nigel, S., & Scrutton, N. S. (Eds.). (2009). *Quantum tunneling in enzyme-catalyzed reactions*. Cambridge: Royal Society of Chemistry (RSC).

Bell, J. (1990). Against "measurement.". *Physics World, 3*, 33–40.

Bohr, N. (1958). *Atomic physics and human knowledge* (p. 20). New York: Wiley.

Born, M. (1969). Symbol and reality. In *Physics in my generation* (pp. 132–146). New York: Springer.

Conrad, M. (1994). Amplification of superpositional effects through electronic-conformational interactions. *Chaos, Solitons & Fractals, 4*, 423–438.

Dehaene, S., & Brannon, E. (Eds.). (2011). *Space, time and number in the brain: Searching for the foundations of mathematical thought*. Oxford: Elsevier.

Delbrück, M. (1949). A physicist looks at biology. *Transactions of the Connecticut Academy of Arts and Sciences, 38*, 173.

Eddington, A. S. (1929). *The nature of the physical world*. Cambridge: The University Press.

Edelman, G. M. (2006). *Second nature: Brain science and human knowledge*. New Haven/London: Yale University Press.

Fröhlich, H. (1975). The extraordinary dielectric properties of biological materials and the action of enzymes. *Proceedings of the National Academy of Sciences, 72*, 4211–4215.

Gell-Mann, M. (1994). *The quark and the jaguar*. New York: W. H. Freeman.

Gerlich, S., Eibenberger, S., Tomandl, M., Nimmrichter, S., Hornberger, K., Fagan, P. J., Tüxen, J., Mayor, M., & Arndt, M. (2011). Quantum interference of large organic molecules. *Nature Communications, 2*, 263. doi:10.1038/ncomms1263.

Greenberg, J. (1978). *Universals of human language*. Stanford, CA: Stanford University Press.

Gross, P. L. K. (1927). The stackability of tetrakaidecahedra. *Science, 66*, 131–132.

Hadamard, J. (1954). *The psychology of invention in the mathematical field*. New York: Dover.

Hameroff, S. R., & Penrose, R. (1996). Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. *Mathematics and Computers in Simulation, 40*(3–4), 453–480.

Harnad, S. (1990). The symbol grounding problem. *Physica D, 42*, 335–346.

Harris, Z. (1991). *A theory of language and information: A mathematical approach*. Oxford: Clarendon.

Hertz, H. (1956). *The principles of mechanics* (pp.1–2) New York: Dover. Original German ed. *Prinzipien Mechanik*, 1894.

Hockett, C. (1960). The origin of speech. *In Scientific American, 203*, 89–97.

Kauffman, S. A. (1993). *The origins of order*. Oxford: The University Press.

Kauffman, S. (2008). *Reinventing the sacred*. New York: Basic Books.

Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from*. New York: Basic Books.

London, F. (1961). *Superfluids* (2nd ed., Vol. I, p. 8). New York: Dover.

McCulloch, W. S. (1988). *Embodiments of mind*. Cambridge, MA: MIT Press.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5*, 115–133.

Monod, J. (1971). *Chance and necessity*. New York: Knopf.

Nairz, O., Arndt, M., & Zeilinger, A. (2003). Quantum interference experiments with large molecules. *American Journal of Physics, 71*(4), 319–325.

Newell, A. (1980). Physical symbol systems. *Cognitive Science, 4*, 135–183.

Ollivier, H., Poulin, D., & Zurek, W. H. (2004). Objective properties from subjective quantum states: Environment as a witness. *Physical Review Letters, 93*, 220401 [4 pages].

Pattee, H. H. (1961). On the origin of macromolecular sequences. *Biophysical Journal, 1,* 683–710. The paper is available from the NIH Archives at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1366362/?tool=pubmed

Pattee. H. H. (1971). Can life explain quantum mechanics? In T. Bastin (ed.) *Quantum theory and beyond* (pp. 307–319). Cambridge : Cambridge University Press http://binghamton.academia.edu/HowardPattee/Papers/1188403/Can_life_explain_quantum_mechanics

Polanyi, M. (1968). Life's irreducible structure. *Science, 160*, 1308.

Pólya, G. (1968). *Mathematics and plausible reasoning: Volume II patterns of plausible inference*. Princeton, NJ: Princeton University Press.

Prashant, & Chakrabarty, I. (2007). Non existence of quantum mechanical self replicating machine. arXiv:quant-ph/0510221v6

Schlosshauer, M. (2006). Experimental motivation and empirical consistency in minimal no-collapse quantum mechanics. *Annals of Physics, 321*, 112–149.

Schmidt, F. O. (1961). Quoted in "concluding remarks,". In A. Sovijärvi & P. Aalto (Eds.), *Proceedings of the 4th international congress of phonetic sciences* (p. xxviii). The Hague: Mouton & Co.

Sereno, M. I. (1991). Four analogies between biological and cultural/linguistic evolution. *Journal of Theoretical Biology, 151*, 467–507.

Simon, H. A. (1962). The architecture of complexity: Hierarchic systems. *Proceedings of the American Philosophical Society,* 467–482. (Reprinted in *The Sciences of the Artificial*, 3rd ed., 1996, Cambridge: MIT Press).

Stent, G. (1966). *Phage and the origin of molecular biology* (p. 4). Plainview: Cold Springs Harbor Laboratory on Quantitative Biology.

Vedral, V. (2011). Living in a quantum world. *Scientific American, 304*(6), 38–43.

von Neumann, J. (1955). *Mathematical foundations of quantum mechanics* (pp. 418–421). Princeton, NJ: Princeton University Press.

von Neumann, J. (1966). In A. W. Burks (Ed.), *Theory of self-reproducing automata* (Lec. 5). Urbana: University of Illinois Press.

Weyl, H. (1949). *Philosophy of mathematics and natural science* (pp. 110–124). Princeton, NJ: Princeton University Press.

Wheeler, J. A. & Zurek, W. H. (Eds.). (1983). *Quantum theory and measurement*. Princeton, NJ: Princeton University Press.

Wigner, E. P. (1960). The unreasonable effectiveness of mathematics in the natural sciences. In *Communications in pure and applied mathematics, vol. 13, No. I*. New York: Wiley.

Wigner, E. P. (1967). The probability of the existence of a self-reproducing unit. In *Symmetries and reflections. Scientific essays*. Bloomington: Indiana University Press.

Wigner, E. P. (1982). The limitations of the validity of present-day physics. In *Mind in nature*. Nobel conference XVII. San Francisco, CA: Harper & Row, p. 119.

Wootters, W., & Zurek, W. (1982). A single quantum cannot be cloned. *Nature, 299*, 802–803.

Zurek, W. H. (1991). Decoherence and the transition form quantum to classical. *Physics Today, 1991*, 36–44.

Part II
The Pattee Papers in Chronological Order

1 The Physical Basis of Coding and Reliability in Biological Evolution

1.1 What Is a Theory of Biology?

Within the intellectual discipline of the physicist there has developed a belief in the existence of general and universal theories of nature, and it is the search for such theories which may be said to guide and justify the intellectual efforts of the physicist as well as the design of most physics experiments. What a physicist means by a ‘good theory’ cannot be exhaustively spelled out. Of course it must include ‘fitting the data’ or ‘predicting observations’ in some general sense. However, much deeper and more obscure criteria are also applied, often tacitly or intuitively, to evaluate the quality of a physical theory. For example, general theories can never be ‘just so’ stories which are only built up bit by bit as data accumulate. General physical theories often stem from relatively simple hypotheses that can be checked by experiment such as the constancy of the speed of light and the discrete energies of photons from atoms, but they must also be founded upon broad principles that express concepts of conservation, invariance, or symmetry. These abstract principles come to be accepted because from our experience we find that in some sense they appear unavoidable. In other words, without such principles it is difficult even to imagine what we mean by a general physical theory of the universe.[1]

Traditionally, in biology, the relation of theory to experiment has been more remote. Much of what is sometimes called biological theory appears to the physicist as a ‘just so’ story, since it is often only a mathematical formalism designed for the practical solution of a specific type of problem and has no direct relation to general physical laws. This situation is often ascribed to basic differences in the subject matter of the physical and life sciences. Perhaps this lack of a basic biological theory is at the root of unresolved historical vitalist-mechanist arguments since much biological terminology never even makes contact with the language of physics.

Reprinted from *Towards a Theoretical Biology, 1, Prolegomena*, C. H. Waddington, Ed. Edinburgh: Edinburgh University Press, 1968, pp 67–93.

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7,
DOI 10.1007/978-94-007-5161-3_2, © Springer Science+Business Media Dordrecht 2012

But recently, following the so-called molecular biological revolution, there have been many statements that now, at last, the mystery of life has indeed been reduced to physical language and laws. In particular we find biochemists and molecular geneticists claiming that they have shown that normal physical and chemical laws provide a relatively clear and simple basis for understanding heredity and most aspects of metabolism. The Watson-Crick template DNA model is commonly accepted as the central concept which is said to reveal the mystery of heredity,[2] and similarly, the detailed structure of proteins has been said to provide a basic understanding of enzyme mechanisms.[3] A common working assumption of molecular biologists is that the remaining problems will be solved by additional experiments. In any case, they do not see any obstacles or essential mysteries on the horizon.[4] This leads to the attitude that biology is explained in terms of ordinary existing physical laws and that therefore no great effort is necessary to apply physical theory to living matter.

On the other hand, in spite of these detailed factual descriptions of polynucleotide and polypeptide interactions in the cell, many physicists as well as biologists remain uneasy. Is this vast amount of phenomenological description really sufficient to support the claim, which is now made even in elementary biology textbooks, that we have a fundamental understanding of living matter in terms of physical laws, that heredity has proven, after all, to be extraordinarily simple, and that the remaining unknowns about living matter are only details to be filled in by more experiment? Can we say with justification that we understand how the laws of physics explain the essential nature of life?

In the remainder of this paper I shall attempt to express why this claim that biology has now been understood in terms of physical laws is not yet convincing. I shall also give some reasons for concluding that the central mysteries of living matter are not to be solved only by collecting more data. Furthermore, I shall propose that even to make a basic distinction between living and nonliving matter some fundamental logical and physical problems remain to be solved at the quantum mechanical level. In particular, I shall argue that any fundamental theory of biology must describe the physical basis of enzymatically controlled hereditary processes that possess the reliability necessary for evolution, and that this will require what amounts to a deeper understanding of the quantum theory of a molecular measurement process.

1.2 Current Molecular Biological Descriptions

There is no need here to repeat in any detail the descriptions used in modern molecular biology, since so many reviews are now available. By molecular biological description we shall mean the use of such concepts as the template replication of DNA, the transcription of the genetic message from DNA to messenger RNA, the translation of this coded message to amino acids, and the synthesis of proteins.[5] An enormous amount of detail is now known about these processes and much more will undoubtedly be discovered in the near future. The principal question, however, does not have to do with the quantity or quality of these data, but rather with their physical interpretation. In particular we want to discuss whether or not these molecular biological descriptions allow us to conclude that the nature of living matter can now be understood in terms of the general laws of physics.

Normally when the physicist says he understands, say, the chemical bond in terms of general physical laws, he does not mean simply that he is optimistic that chemical bonds are consistent with quantum mechanics, or that if he cared to go into the matter he would find no serious problem in describing the chemical bond by the rules of quantum mechanics. On the contrary, although the chemical bond was first recognized and discussed at great length in classical terms, most physicists regarded the nature of the chemical bond as a profound mystery until Heitler and London quantitatively derived the exchange interaction and showed that this quantum mechanical behaviour accounted for the observed properties of valency and stability. On the other hand, it is not uncommon to find molecular biologists using a classical description of DNA replication and coding to justify the statement that living cells obey the laws of physics without ever once putting down a law of physics or showing quantitatively how these laws are obeyed by these processes. Of course, as a speculative prediction such statements are acceptable. But certainly nothing could be less fruitful than allowing this most fundamental and challenging question of whether living matter can be reduced to the basic laws of physics to be obscured by such pronouncements from molecular biologists without some regard for the established language and laws of physics.

1.3 What Is the Question?

Let us for the moment assume that the experiments of molecular biology and genetics have indeed shown that no detailed process in living matter evades or violates normal laws of physics. If this were the case, does the question of the nature of life appear answered? In other words, even if it were the case that living matter was exactly the same as nonliving matter with respect to description by physical laws, would we then say that we fully understand life in terms of physical laws? No, I think not, because this does not answer the obvious question of why living matter is so conspicuously *different from* nonliving matter. In other words, we do not find the physical similarity of living and nonliving matter as puzzling as the observable differences. Before we can attempt to explain these differences in terms of physical laws we must state clearly what these differences are. Older biology texts usually begin by listing the ‘characteristics of life’ that may include growth, reproduction, irritability, metabolism, etc., but these are not general enough concepts. What is the most general property of life that distinguishes it from nonliving matter? Certainly the most *general* property is the potential to evolve. Therefore the fundamental question can be restated: Is the process of biological evolution understandable in terms of basic laws of physics?

1.4 Two Basic Assumptions

In order to show the difficulties in answering this question let us restate the situation in the form of assumptions:

Assumption A. Living states and nonliving states of matter are in no way distinguishable by their detailed description in terms of initial conditions or elementary laws of motion. i.e. both living and nonliving forms of matter obey precisely the same physical laws.

Assumption B. Living states of matter are distinguishable from nonliving states of matter only by the potential for evolution. i.e. the hereditary transmission of naturally selected traits.

To make these assumptions more plausible let us consider for a moment the antithetical assumptions. Suppose, for example, that the difference between living and nonliving matter depended upon different initial conditions. From the point of view of the physicist we would have to call this a 'special creation' that may be allowable as a highly unlikely event or a miracle; but this would nevertheless be scientifically barren since it can be neither derived from any physical theory nor tested by any real experiment.[6] Furthermore, if we assumed that living and nonliving matter obey different elementary laws of motion, then by the physicist's meaning of a law, there must be observable or derivable regularities or correlations between detailed measurements involving one type of matter but not the other. Since an enormous number of observations have been made and no such regularities have been found, this antithetical assumption seems unjustified. Notice that Assumption A does not imply that all aspects of physical theory have been formulated, but only that whatever theories we currently accept must apply equally to living as well as nonliving matter. Finally, if we reject Assumption B and assume antithetically that living and nonliving matter can both evolve in some sense, then we have only succeeded in generating a new question: Why did living matter distinguish itself by evolving so much more variety than nonliving matter? In other words, we must have in addition to Assumption A, which states the *similarity* of nonliving and living matter, a second assumption which clearly *distinguishes* living from nonliving matter. To omit the second type of assumption is to miss the whole problem.

Accepting Assumptions A and B for our discussion, what can we conclude from them? Some physicists feel that such assumptions are contradictory. Wigner's[7] argument that self-replication is impossible, assuming only the normal laws of quantum mechanics, would fall into this category. Other physicists propose that autonomous biological laws must exist. Such arguments have been given by Bohr,[8] Elsasser,[9] and Burgers,[10] for example.

My own point of view is that there is no scientific value whatever in attempts to dismiss such arguments because they have their basis more in the language or logic of physics rather than in the details of molecular biology. Assumptions A and B are statements of a crucial paradox which must be zealously and carefully pursued if we are to have a physical theory of general biology. Furthermore, I believe there is reason to expect that these assumptions are closely related to the central epistemological paradox of the mind-body problem itself. However, in this paper I shall emphasize this paradox only in the context of the origin of life problem. First, I shall try to clarify these assumptions so as to sharpen the paradox. Otherwise the central problem can too easily become obscured by the many details of new experimental discoveries.

1.5 What Are the Physical Laws?

The Assumption A is relatively easy to amplify because the meaning of initial conditions and laws of physics have already been deeply analyzed.[11] What we wish to emphasize, however, is that the physicists meaning of ‘obeying the laws of motion’ is a rigorous statement which can be quantitatively verified by measurement and calculation. An elementary law of motion is a prescription for correlating the values of certain variables which give the state of a system at any one time to the values of these variables for any other time. In this language, once the complete state of a given system has been chosen by assigning initial conditions for one time, any additional information about an earlier or later state of the system is redundant. That is, no better prediction about the future or past of the system can be made, in principle, by supplying more information. The rules for applying this descriptive language are precisely formulated and one cannot, for example, say that a molecule obeys these elementary dynamical laws of physics simply by looking at numbers representing the *average* structure of a large collection of these molecules or by moving around a desk-top classical model of one of these molecules. In other words, to say that an enzyme or nucleic acid molecule obeys the dynamical equation of motion of quantum mechanics cannot be regarded by the physicist as a justifiable conclusion without some evidence to actually support such statements.

We have therefore labeled our statement A as an assumption, because although it might be argued that quantum mechanics has in the past described correctly many diverse molecular effects, we must also consider the arguments that have been presented showing that quantum mechanics is not consistent with the basic property of self-replication.

In the clarification of Assumption B we encounter another type of difficulty. Few biologists would dispute that the living states of matter evolve by a different process than the nonliving states. In fact, the potential for hereditary evolution may be used as a definition of present life. But it might be argued that hereditary evolution is not the most elementary or fundamental condition for the origin of life. For example, simple autocatalysis, metabolism, or replicating processes may also be called primeval features of the living state. However, to be brief, I shall simply define as a necessary condition for the origin and persistence of life the property of reliable hereditary transmission of naturally selected traits. Unfortunately this phrase is not yet in the language of physics, and its meaning is often imprecise even in biology. Therefore let us try to define what hereditary transmission and natural selection can mean in the language of physics.

1.6 What Is Heredity?

The traditional idea of a hereditary process involves the transmission from parent to offspring of particular traits. By a trait the biologist does not mean an invariant of the equations of motion, but one property chosen from a set of possible alternatives. The trait which is actually transmitted depends upon a *description* of the trait recorded or remembered from some earlier time, Thus, the central biological aspect of hereditary evolution is that the process of natural selection operates on the actual traits or phenotypes and not on the particular description of this phenotype in the memory storage which is usually called the gene. This is essential biologically because it allows the internal description or memory to exist as a kind of virtual state which is isolated for a finite lifetime, usually at least the generation time, from the direct interaction which the phenotype must continuously face.

The crucial logical point of hereditary propagation which corresponds to the biological distinction between genotype and phenotype is that hereditary propagation involves a *description* or *code* and therefore must require a *classification* of alternatives and not simply the operation of the inexorable physical laws of motion on a set of initial conditions. As we stated in the last section, these laws of motion tell us how to transform the state of a system at a given time into the state at any other time in a unique and definite way. The equations of motion are therefore said to perform a one-to-one mapping, or more specifically, a group transformation of the states of a system. On the other hand, the hereditary process that must transmit a particular trait from a larger set of alternatives must perform a classification process, and this involves a many-to-one mapping. It is for this reason that concepts such as memory, description, and code that are fundamental in hereditary language are not directly expressible in terms of elementary physical laws. Direct copying processes, such as crystal growth or complementary base pairing in DNA do not involve a code or classification of alternatives; and therefore, even in classical language, simple template copying processes are not a sufficient condition for evolution by natural selection. When there is no distinction between genotype and phenotype or between the description of a trait and the trait itself or, in other words, when there is no coding, process which connects the description by a many-to-one mapping with what is described, then there can be no process of hereditary evolution by natural selection.

The logical aspects of this fundamental evolutionary principle were understood by von Neumann[12] in his design of a self-replicating automaton based on the Turing machine. It is significant that von Neumann's self-replicating automaton has the same basic logic that is now known to exist in cells, even though his replicating automaton was designed without any knowledge of the details of the cellular translation code and the roles of nucleic acids and enzymes. Nevertheless it was clear to von Neumann that simple template replication or copying in itself was of no interest in either the logical or the evolutionary sense, and that only a concept of heredity which includes a code could provide growth of complexity that had any real significance for learning and evolution. Thus it may be said that a threshold of logical complexity exists for the origin of evolving hereditary structures. Following von Neumann's work many papers have pursued the interesting and essential logic of this problem.[13] It is remarkable how few biologists are aware of this work and of the logical basis for a coding process in hereditary biological evolution.

1.7 The Central Problem

We have now given some idea why the elementary laws of physics do not seem directly suitable for describing hereditary behavior. At the logical level we may say that the laws of physics describe a one-to-one mapping process, whereas hereditary propagation requires a many-to-one mapping process. Or in more physical terms we may say that the elementary physical laws are symmetric with respect to time, whereas hereditary propagation requires a direction to time. Or in other words, the temporal relation between the memory of a trait and a trait itself is not symmetric.

There is of course a broad general theory of physics called thermodynamics which is capable of treating irreversible phenomena. We may therefore ask if thermodynamic or statistical mechanical theories cannot be applied to hereditary phenomena. The answer is that of course they can be applied, but they do not lead us to expect biological evolution. In fact, it is the second law of thermodynamics that at first sight appears to be the antithesis of biological evolution leading as it does to complete disorder as opposed to the increasing complexity of biological organisms. We may therefore say that the problem of describing hereditary processes in terms of the laws of physics must not only overcome the difficulty in deriving irreversible phenomena from reversible laws, but in addition it must also show how the consequences of hereditary irreversibility lead to the phenomenon of evolution in living matter rather than the complete thermodynamic equilibrium of nonliving matter.

1.8 The Classical Evasion of the Central Problem

One popular concept of living matter that seems to evade this paradox is the so-called automata description of molecular biology. This description treats the cell as a classical machine that behaves very much like a modern large-scale computer.[14] Such classical machines clearly exhibit the property of memory storage and hereditary transmission as well as coding and classification processes. How are such classical machines described in terms of the laws of physics?

This can be done only by the introduction of a certain type of structure that controls to some extent the dynamic motion of the system, but which is not derivable directly from the basic equations of motion. In order to exhibit the fundamental hereditary property of classification, or the selection of a trait from a larger set of alternative traits, there must be available more degrees of freedom in the static description of the machine than are available for the dynamic motion of the machine.

In other words, the very concept of a memory in a hereditary system implies the existence of more freedom in the static state description than in the motion of the system, since it must be dynamically constrained so as to propagate only that particular trait that is recorded in the memory storage. Such a structure that has more degrees of freedom in its state description than in its dynamic motion is called in classical physics a *non-holonomic* constraint.[15] If one accepts the classical description of non-holonomic constraints, it is possible to tailor a machine to represent almost any code or logical function that one can imagine, and this is the basis of all computer design. In fact, it is possible to program large-scale digital computers to imitate macromolecular processes in living cells, including DNA replication, transcription, and coding into protein enzymes.[16] We therefore must raise the question: Are classical descriptions or models of living cells an adequate basis for understanding the fundamental nature of living matter and its evolution?

A part of the answer to this question was already suggested by the physicist Schrödinger[17] in his book *What is Life?* that appeared in 1944. Schrödinger pointed out that the order that we associate with classical mechanisms is based on the averages of large numbers of molecules, whereas the order in the cell is based on single molecules. Schrödinger suggested that the relative stability of individual molecules can be understood in terms of the stationary states of quantum mechanical systems, but he did not discuss the transmission of this order into macroscopic systems, that is, the expression of this order as a hereditary trait. This is another statement of the central problem that still must be solved.

In order to present the problem in more detail, let us return to the classical concept of a hereditary system which must involve a non-holonomic constraint. What are some of the basic properties of non-holonomic systems? The idea of a constraint is entirely classical, arising from the treatment of some degrees of freedom as purely geometrical structures which do not depend on time and the laws of motion. However, when we look at matter in more detail, we realize that all macroscopic structures must ultimately be represented by elementary forces, which hold them together. We may then distinguish permanent structures as only metastable configurations with relatively long relaxation times compared to our time of observation. For example, an ordinary clock which may, during short intervals, appear to be telling very accurate time will, over longer intervals, slowly lose this accuracy and gradually approach irreversibly the equilibrium to which all classical machines must tend. A good clock is simply a mechanical device that manages to measure the same time interval a large number of times before it reaches equilibrium. Thus at least two widely differing relaxation time scales are necessary for the description of hereditary behaviour in statistical systems, and at least one of these time scales must describe an irreversible process. Usually one of these time scales is so long that it is neglected in the treatment of the dynamical problem, and it is replaced only by geometric constraints. The more complete mathematical description of this classical hereditary behaviour in non-equilibrium, non-linear statistical mechanical systems can become very elaborate.[18] But, as Schrödinger pointed out in the case of hereditary storage, the peculiarity of biological chemistry is that all its hereditary processes are based on the dynamics of individual molecules and not on statistical averages of vast numbers of molecules. Therefore we must try to extend these classical and statistical mechanical ideas of a hereditary process to individual reactions at the quantum mechanical level.

But in view of the obvious difficulty of such a microscopic description we may again raise the question: Why is it necessary to use quantum mechanical description when it is known that in many cases, even in chemistry, a classical description is adequate for a good understanding of the processes involved? In other words, why is it not possible to admit that a quantum mechanical description would indeed be more accurate, but that for all practical purposes a classical description is close enough?

1.9 The Reliability Condition for Evolution

Now we have asked the crucial question: When is a theory or a description ‘close enough’? We have asked this question about our own attempts at describing living matter in terms of physical laws; but certainly the same question can be applied to the hereditary process itself, and we may ask: When is the description of a hereditary trait ‘close enough’? This is a very practical type of question, and its answer depends upon what purpose one has in mind for a particular theory or hereditary description. In the context of the origin of life we may restate this question as follows: When is hereditary storage and transmission reliable enough to achieve the persistent evolution of complexity in the face of thermodynamic errors, that is, in the face of the second law of thermodynamics? Even though we do not understand the mechanism, the only conclusion I have been able to justify is that living matter has distinguished itself from nonliving matter by its ability to achieve greater reliability in its molecular hereditary storage and transmission processes than is obtainable in any thermodynamic or classical system.

Now while it is reasonable to assume that the relatively high reliability of hereditary *storage* in cells is based upon the quantum mechanical stationary states of single molecules, we must still find an explanation for the relatively high reliability of the *expression* of these hereditary descriptions as classical traits that interact with the classical environment. In other words, we may say that the description of the trait is quantum mechanical, whereas the natural selection takes place on the classical level between the phenotype and the environment. But even though we do not understand the hereditary transmission process, the answer to our question whether classical laws are ‘close enough’ for a theory of life is now obvious; for if the cell itself cannot use a classical description for its hereditary processes, then how could we expect to describe this unique biological reliability only in terms of classical description?

We must next ask what type of physical theory can be used to describe the expression of a quantum mechanical hereditary *description* as classical interactions between the phenotype and the environment. In particular, by what physical theory do we describe the hereditary transmission process that decodes the quantum mechanical description to produce the classical phenotypic expression?

1.10 The Quantum Theory of Measurement

There are a few other types of phenomena in physics in which quantum and classical descriptions must be closely related — ferromagnetism, low-temperature phenomena such as superconductivity and superfluidity, and the measurement process in quantum mechanical systems. It is significant that for all these types of phenomena there exists no complete description in terms of elementary quantum mechanical equations of motion. For this reason, while it does not appear likely that an explanation of molecular hereditary transmission will be produced forthwith, at least the problem is not entirely foreign to physics. Therefore while I cannot support the optimistic belief of many molecular biologists that heredity is simple and has now been explained in terms of physics, neither can I be as pessimistic as some physicists in their assertion that living states of matter cannot be derived from physical laws.

The problem of describing a measurement process in terms of the quantum equations of motion has evaded clarification since the formulation of quantum mechanics. Since there are many papers which discuss the problem in detail,[19] I shall do no more here than suggest how molecular hereditary processes are related to the quantum theory of measurement. The basic problem may be stated in the following way: The quantum equations of motion operate on unobservable wave functions that may be interpreted as probability amplitudes. Under certain conditions, these unobservable probability amplitudes can be correlated with observable variables in the normal classical world, and when this happens we can say that a quantum mechanical measurement has been executed. However, the quantum equations of motion do not appear to account for this correlation of probability amplitudes with the observable probabilities in the classical world, and a second type of transformation called 'the reduction of the wave function' must be used to produce a measurable quantity. The quantum equations of motion are reversible in time and perform a one-to-one transformation of the wave functions, whereas the reduction of the wave function or measurement is an irreversible process and involves a classification of alternatives or a many-to-one transformation. This necessity for two modes of description is at the root of the wave-particle duality, the uncertainty principle, and the idea of the necessity of complementarity in the complete description of quantum events.

However, it is also this duality that leads to the conceptual difficulties of measurement processes, since there is as yet no objective procedure for specifying where in a chain of events a measurement occurs. In other words, whether or not a measurement is said to occur depends somewhat arbitrarily on where the observer chooses to separate his quantum mechanical and classical descriptions of a given measurement situation. If he chooses to consider the entire system, including what he would normally call the measuring instrument, as only a single quantum mechanical system, then he could recognize no measurement. In the same way, if he chooses to treat a collection of molecules which includes what he normally would call a hereditary memory as only a single quantum mechanical system, then he could recognize no hereditary process.[20]

1.11 Enzymes as Measuring Molecules

In view of the unsatisfactory state of the theory of measurement in quantum mechanics, it is a remarkable fact that physicists continue to make accurate measurements, just as biologists continue to replicate, without in a sense, understanding what they are doing. However, in the case of physicists this can be partially explained by the size of measuring devices, which are usually large enough to be clearly recognized and treated only as classical systems. In any case, measuring devices are designed by men and are not considered as spontaneous collections of matter. On the other hand, we cannot make this excuse for biological replication. When we speak of individual molecular hereditary transmission as similar to a measurement process, we must ask what corresponds to the measuring instrument at this microscopic level. Or in terms of the origin of life, what is the simplest molecular configuration which could express a hereditary trait and which we could have expected as a reasonable spontaneous molecular organization?

Here we must return to our fundamental definition of heredity as a classification process rather than as simple copying, or the propagation of an invariant of the motion. We have pointed out that a classical physical representation of a classification process must depend on non-holonomic constraints, that is, on structures which allow more degrees of freedom in the state description than is available for the actual dynamic motion of the system. At the molecular level this would imply that non-holonomic constraints allow a larger number of energetically possible reactions than the number of reactions which are actually available to the dynamics of the system. Now in chemical terms, reactions which are *available* as distinct from those which are energetically possible can differ only in the activation energy and entropy, so that we are led to associate the classification process or hereditary propagation with the control of rates of specific types of chemical reactions. Of course, in cells the control of rates and specificity is accomplished by the enzyme molecules. Furthermore, it is significant that classical models of enzyme mechanisms depend upon flexible structures or allosteric[21] and induced-fit[22] descriptions that are equivalent to the physicists' non-holonomic constraints. It is of course possible that other molecules such as nucleic acids also exhibit nonholonomic, catalytic properties, but this remains to be demonstrated.

As we have already noted, the physicist may design and perform experiments on quantum mechanical systems without microscopic analysis of the process of measurement since in most cases a distinction between the quantum system being measured and the classical measuring device can be clearly specified or recognized. In other words, we accept the non-holonomic constraints of a clock, a switch, or gate mechanism because these are large classical devices with many degrees of freedom which we can statistically tailor to approximate our needs with the desired precision or reliability. But at the microscopic level it is by no means obvious that we could design a single molecule which performs with the speed and reliability observed for specific enzyme-controlled reactions. In the first place, the very idea of a non-holonomic constraint in an elementary quantum mechanical system forces on us a profound modification of the language.[23] Not only would the idea of measurement have to be extended to include non-observed quantities, but also the equations of motion are effectively modified by non-holonomic conditions, since there is no possibility in deriving such exact constraints by taking into account additional existing degrees of freedom. On the other hand, this requirement of a reliable microscopic non-holonomic constraint is consistent with the early suggestion of London,[24] and more recent suggestion of Little,[25] that macromolecules could conceivably possess superfluid or superconductive states which would allow change of shape or transfer of matter with no dissipation. As London pointed out such a quantum fluid state would combine the characteristic stability of stationary states with the possibility of dynamic motion isolated from thermal agitation. This is precisely what would appear to be essential for specific catalysts which act as precise molecular measuring devices.

A direct experimental test of such a measurement theory of specific catalysis may run into a type of difficulty foreseen by Bohr, namely that external measurements of crucial life processes may be incompatible with the results of the process. If measurements by single enzyme molecules depend upon the internal correlation of their electrons, then any device which can be said to perform an external measurement on these electrons will necessarily destroy some of these correlations with the result that specificity and catalytic power of the enzyme will be correspondingly decreased. However it is not clear that other more indirect evidence may not be obtained to test such a theory.[26]

It is to be expected, of course, that classical description will indeed be useful at many points, and that for many practical applications the details of the quantum mechanical description are unnecessary. However, in terms of any general theory of biological systems the *reliability* of hereditary transmission or the speed and accuracy of measurement is crucial. For example, the difference between a mutation rate of 10^{-4} and 10^{-8} per elementary hereditary transmission may easily be the difference between the immediate extinction or long evolution of a species, and no one could claim that this is a trivial difference.[27] It is this quantitative difference in the speed and reliability of hereditary transmission for which quantum mechanics can account and for which classical theory cannot.

In terms of the origin of life problem, this assumption also leads us to believe that *life began with a catalytic coding process at the individual molecular level,* since no spontaneous thermodynamic system or classical machine appears to provide the necessary speed and reliability for such a distinctive evolutionary process within the classical environment. Therefore, although with great effort we may design complicated classical hereditary machines which may adapt themselves to a classical environment for a limited time, we would not expect such complex devices to arise spontaneously on the primitive earth, nor could we expect them to achieve a statistical reliability in their hereditary processes which would allow them to distinguish themselves so successfully from the environment for five billion years.

1.12 Design of Origin of Life Experiments

What type of abiological experiments does this measurement theory of hereditary processes suggest? First of all we are led to believe that specific catalytic molecules are essential for the coding process in hereditary transmission. Contrary to the so-called central dogma, which states that nucleic acids transmit all hereditary information and that proteins can only receive it, we would have to conclude that while template molecules or holonomic structures may be said to *store* hereditary information it is only the non-holonomic or allosteric catalysts that can *transmit* hereditary information. Moreover, it is important to realize that *a definition of stored information itself cannot usefully be made without a complete specification of the coding mechanism for transmitting it.* Without complete specification of the transmission code there is no way to determine what variables of a given physical structure consist of hereditary information which is to be transmitted, and what variables are simply to be treated as initial conditions needed to specify the storage structure at a given time. Failure to recognize that prior specification of the transmission code is necessary in order to define stored information in an objective way has led to much confusion in the use of the information concept particularly in biological systems.

The experimental approach suggested by this theory contrasts sharply with the strategy of most so-called ‘chemical evolution’ or abiogenic organic synthesis experiments which emphasize the growth of non-hereditary chemical complexity as judged by the similarity of particular spontaneous species of molecule with existing biochemical species in cells.[28] While it may be relatively easy to compare the similarity of these spontaneous molecules with the evolved molecules of cells, the question of the significance of each type of molecule is left open. This has generated much discussion as to which type of synthesis is most closely related to the origin of life on earth and elsewhere. Since widely different sets of initial conditions can produce many of the same organic molecules, there have also arisen controversies over such uncertainties as the equilibrium conditions and free energy sources which actually produced the first prebiological molecules on the earth, and what extraterrestrial conditions might favour the occurrence of certain types of prebiological molecules.

I would like to point out that from the hereditary point of view it makes little difference for the general origin of life problem whether a molecule is made by heat ultraviolet ionizing particles, or for that matter obtained from a chemical supply house, *as long as the molecule has no memory.* Furthermore, since we can associate hereditary transmission only with rate control processes, or in other words, since equilibrium states can have no memory, we should not expect equilibrium conditions to play a primary role in the origin of life. Of course I do not mean that organic syntheses and equilibrium considerations are not important for the origin of life problem. What I wish to emphasize is that the hereditary property itself is the only context from which these other questions can have any objective biological interpretation. Our theory therefore constrains us to look for the simplest possible hereditary chemical reaction *processes* before we can usefully compare our chemical products with living cells.

1.13 Examples of Hereditary Copolymer Reactions

How shall we experimentally recognize the most primitive hereditary reactions or codes in simple molecules? This is a very difficult question which I cannot fully define, but the general idea can be illustrated by a series of examples of polymer growth. Consider first a simple growing homopolymer in which there is an initial monomer addition rate constant, K_a. After the chain grows long enough, suppose that it folds into a helical conformation, say, with five monomers per turn, and that because of the folding the monomer addition rate increases to $K'_a > K_a$. The nature of the bond is not changed, only the rate has increased. One case of such conformation-dependent catalysis occurs in the N-carboxyanhydride synthesis of polypeptides.[29] The significant aspect of this simple conformation-dependent, rate-controlled reaction is that the oldest exposed monomer in a helical chain is controlling the rate of addition of the next monomer. This amounts to a delay in the control mechanism corresponding to one turn in the helix. Now this *delayed control process* may not appear to have much evolutionary potential. However, we shall show how natural modifications of such *conformation-dependent specific catalytic effects* may produce elaborate hereditary coding in simple copolymers.

Next consider a copolymer growth in which the initial comonomer addition rates are K_a and K_b. Suppose that this chain also folds into a helix with five monomers per turn and that in this configuration the proximity of the (n − 4)th to the (n + 1)st position catalyses the next addition step as in the previous example. However, now when we are using two types of monomer it is generally unlikely that the catalytic effect of the (n − 4)th position is independent of the type of monomer at that position. If we now assume that there is a very strong rate controlling effect of only the (n − 4)th monomer on the addition of the next monomer, there will then be four possible control schemes or *codes* as shown in Table 1.1.

What will be the effect of these possible codes on the sequences in the copolymer chain? The last two codes will clearly degenerate into simple homopolymers no matter what the starting sequence may be. However, the first two codes will lead, respectively, to eight and four species of periodic copolymer. It is also clear that the linear sequence in each of these species is completely determined for a given code

Table 1.1 The four possible codes generating copolymer sequences

Code	Monomer types in (n − 4)th position	Catalysed monomer type in (n + 1)st position
1	a	a
	b	b
2	a	b
	b	a
3	a	a
	b	a
4	a	b
	b	b

of Table 1.1 by any five adjacent monomers in a helical turn. and therefore each turn of the helix can be considered as a genetic sequence. For example, if an *a* or a *b* monomer at the (n – 4)th position increases the relative rate of addition of the same type of monomer as shown in the first code of Table 1.1, then any of the five cyclic permutation sequences, a*baba, babaa, abaab, baaba,* and *aabab* are equivalent genetic sequences for one of the species. The other seven species are generated from the two homopolymers, *aaaaa* and *bbbbb,* and the sequences *babab, aabaa, bbaba, baaab,* and *abbba* or one of their cyclic permutations. It is important to realize that the specificity or relative catalytic power of the (n – 4)th, monomer, or in other words the *reliability* of the tactic catalyst with respect to the types of added monomer will determine the inherent rate of mutation in this type of hereditary propagation. Of course, the addition of an uncatalysed monomer, that is, the addition of a non-coded monomer, will not necessarily lead to a new species, since all cyclic permutations of the end-turn sequence are genetically redundant. This would correspond to a mutation in DNA that still codes for the same amino acid.

Suppose now that we wish to increase the reliability of such a coding process. In other words, we wish to increase the specificity and corresponding catalytic power for the addition of particular monomers. One reasonable mechanism for accomplishing this is to assume that more monomers must play a role at the active site, or in other words, that there are more interactions with the monomer which is to be added. Using the same basic model of a helical copolymer, suppose that not only the (n – 4)th position monomer determines the type of addition but that the last monomer or nth position also influences the specificity. This is sterically reasonable, since the nth and the (n – 4)th monomer form a step dislocation in the helix at the position where the next monomer will be added. But now instead of only 4 possible coding schemes as shown in Table 1.1 there are 16 possible codes, again assuming only absolute specificity or so-called eutactic control. If we choose the code which catalyses the addition of an a-type monomer when the nth and (n – 4)th monomer are the same type and a b-type monomer when the nth and (n – 4)th monomer are a different type, we will obtain four species of copolymer which may be represented by the four periodic sequences given below:

S_1: $(a)_n$
S_2: $(bba)_n$
S_3: $(bbbaaba)_n$
S_4: $(bbbbbababaabbaaabaaaa)_n$

The molecules within each species S_2, S_3, and S_4 will differ from each other only in the phase of the starting sequence. The sum of the length of all periods is $2^5 = 32$, and therefore no other eutactic species are possible for this given conformation and code. Of course, we may also specify each species by five consecutive monomers from any part of each chain. For example, S_1: *aaaaa*, S_2: *abbab,* S_3: *baaba*, S_4: *bbbbb.* It is clear that species S_2, S_3, and S_4 have 2, 6, and 20 other equally good starting genetic pentamer sequences respectively.

If one forms a state-transition matrix for this polymer growth process listing all 32 initial and final states, the hereditary property will be apparent by the reducibility of this matrix into four sub-matrices corresponding to the four species of the chain. From this state-transition matrix description it will be obvious that the growth space for a given initial five-monomer chain is less than the physically possible state space for the five-monomer chains. The mechanism for this growth process, which we have not specified here, is therefore equivalent to a nonholonomic constraint.

Of course, these simplified copolymer models are only to illustrate in the simplest way how true hereditary processes can arise at the molecular level. It is unlikely that tactic polypeptide growth would occur under so few constraints or in this particular autonomous form. The optimum conditions under which such tactic catalytic growth of polypeptides might be found on the sterile primitive earth need further discussion.[30] It is plausible from the known tactic processes in present cells, and the assumption of continuity in evolution, that the most primitive polypeptide tactic catalysis also involved polynucleotides and the constraints of particle or membrane-like surfaces. The origin of the nucleotide-amino acid code remains a deep mystery, but from what we have said, the answer should not be expected in template models or non-catalytic processes.

1.14 The Reliability of Copolymer Catalysts

Even though we are not able to propose at present any detailed quantum mechanical mechanism for this type of conformation-dependent catalytic process, it is instructive to look for specific properties of such single copolymer hereditary catalysts which affect their reliability, since this property is essential for evolution. The significant characteristic of enzyme catalysis is that the specificity may be controlled only by weak bond interaction, whereas the catalysis or rate control operates only on the strong covalent bonds of the substrate. By contrast classical machines, like clocks, use the strong bonded structures, such as the gears and escapements, to control the formation of weak bonds, that is, the frictional contacts between escapement pins and gear teeth. At the copolymer level a distinction between strong and weak bonds is already implicit in the concepts of monomer *sequence* and *conformation*, since neither of these terms could be usefully defined if only one type of bond strength existed between monomers. The linear sequence is in fact defined as the monomer order obtained by following the strong bonds from one end of the chain to the other, while the conformation in linear chains refers to the shapes held by the weak bonds as allowed by the rotation or flexibility of the strong bonds, but not by breaking strong bonds. Of course in enzymes there are covalent bonds cross-linking the chain, but the definition of a linear sequence is still recognized by the most stable strong bond path.

What is the effect of these different roles of strong and weak bond interactions on the reliability of hereditary propagation in classical and quantum mechanical systems? We have already pointed out, following Schrödinger, that the covalent bond in a copolymer chain provides an ideal static *storage* mechanism for hereditary information. However, it is no less important that all dynamic hereditary *transmission* processes, which include replication, transcription, and coding, operate with high reliability in the face of external and internal perturbations. In particular, it is more important that hereditary propagation cease altogether rather than propagate errors or lose the coding rules. Otherwise such uncontrolled catalytic activity only speeds up the destruction of the hereditary information. For example, in the helical copolymer model in which the helical structure is maintained only by weak bonds and the genetic memory by strong bonds we could expect some form of error prevention upon heating, since the helix will become a random coil first and thereby stop catalyzing monomer addition. On the other hand, in most classical machines such as clocks, it is more likely that upon gradual rising temperature the machine will begin to operate with errors before it stops altogether. In other words, unless special error-correcting devices are employed, a classical clock will tell the wrong time before it melts, whereas an enzyme will melt (denature) before it catalyses the wrong reaction. For these reasons we may expect optimum reliability and survival value in hereditary systems in which the non-holonomic constraints representing the translation code mechanism are formed from weak-bonded structures, while the memory storage as well as the phenotypic expression of this description is preserved in strong-bonded metastable structures. Evidence of thermally inactivated specific catalysts should therefore be assigned high significance in abiogenic experiments.

However, even under optimum operating conditions there remains a certain level of random thermal disturbance which affects the speed and accuracy of any classical measuring device. Normally, when Brownian motion or particle statistical fluctuations disturb the accuracy of a measurement the only remedy is to increase the mass of the device or increase the time of observation so as to average out the fluctuations. Consequently high accuracy or precision in classical machines is incompatible with both small size and high rates of operation. We are left then with the challenging problem of interpreting the enormous speed and precision of individual enzyme molecules without being able to use the statistics of the large numbers of degrees of freedom that we associate with macroscopic objects.

At first sight such speed and accuracy in single quantum mechanical systems may appear even more difficult to explain because of the uncertainty principle. For example, we may say that if we choose to measure the energy of a system with an accuracy of ΔE, then the measurement interaction must extend over a time interval of $\Delta t \geq h/\Delta E$ so that speed and accuracy in this case are fundamentally incompatible. However, a more precise description of what enzymes actually accomplish does not involve such a simple relation between conjugate variables involved in the measurements. The specificity of enzymes appears to depend on the accurate fitting of a part of the substrate to a part of the enzyme. This implies that specificity depends on the measurement of relative position coordinates of certain regions of the substrate. But since the bond that is catalyzed may be at a different location, the momentum coordinates conjugate to the coordinates determining the specificity need have no direct relation to the speed of catalysis. On the other hand, if the enzyme structure has non-holonomic properties, which we claim is necessary for hereditary transmission, this implies that dynamic correlations must exist between the measured coordinates determining specificity and the momentum coordinates involved in the catalysis.

The reliability of substrate recognition and the speed of catalysis now become a problem of describing how such dynamical correlations can be maintained without invoking classical structures. As we indicated above, this is a difficult conceptual and mathematical problem.

Such reliability consideration will probably be crucially related to the size of enzymes and the structures associated with hereditary transmission, which of course includes the machinery for DNA replication and transcription as well as coding. It has been shown that the allowable accuracy of quantum mechanical measurements increases with the size of the measuring device, so that only in the classical limit can these measurements be described as exact.[31] This inaccuracy cannot be interpreted as the normal errors of measurement, or associated with the uncertainty of measuring a *pair* of non-commuting variables. Rather it is the result of the attempt to describe the measurement transformation by the quantum equations of motion. Although quantitative estimates of reliability have not been made, it is plausible that copolymers must have grown spontaneously to a certain size before they could perform tactic catalysis with sufficient reliability to assure some evolutionary success. Perhaps such reliability requires membrane or particle-bound copolymers as found in the tactic reactions in present cells.

The main point of this discussion is to emphasize the necessity of reliable molecular coding for any persistent hereditary evolution. There are two aspects to this necessity: first, the *logical threshold* as illustrated by von Neumann (see pp. 75–77) that distinguishes the description or genotype from the construction of phenotype; and second, the *physical reliability threshold* which maintains the hereditary dynamics so that the rate of accumulation of information by natural selection can exceed the rate of error in the overall hereditary transmission process. These discussions suggest that neither template copying processes nor non-specific catalysis can account for the origin of life. Even though classical automata may be designed by man to satisfy the logical and reliability thresholds useful for a kind of hereditary evolution, we would expect that quantum mechanical description will turn out to be essential for any fundamental understanding of living matter.[32] Furthermore, the difficulties in quantum mechanical description of reliable hereditary processes do not appear to be simply a matter of complexity, but are likely to involve some of the most difficult conceptual problems that lie at the basis of physical theory. Would it be so surprising, after all, if the secret of life turned out to be based on something more than simple chemical description?

1.15 Some Broader Questions

I have used the origin of life context in discussing coding and reliability because this level allows the simplest possible conception of a molecular hereditary transmission process. We have seen that even at this level the theoretical difficulties remain serious. Nevertheless I believe that the concepts of coding and reliability will not only be useful, but also crucial at all levels of biological organization—cellular, developmental, evolutionary, and certainly in the higher nervous activity associated with the brain. We have used code to mean the relation between an elementary genotype and a phenotype, that is, a relation between a physical symbolic description and the physical object that is actually constructed from this symbolic description.

The process of cellular replication and in particular the development of the organism may be interpreted as an entire system construction process which requires a coding mechanism which interprets as well as replicates a description. Largely from studying the logic of abstract automata we may begin to appreciate how, through the discovery of simple codes, it is possible to generate elaborate ordered structure from relatively concise descriptions. Such a description-code-construction process cannot be adequately characterized as either preformation or epigenesis, since on the one hand the construction may be totally unlike its description, whereas on the other hand the description and code structure together provide a complete, autonomous generation of the phenotypic construction within the crucial limits of reliability.

At the evolutionary level this concept of a symbolic genetic description and its code structures must be broadened to a larger system that includes not only the description of the system itself but also a description or a ‘theory’ of the environment. In the evolutionary context the phenotype itself now plays the role of a composite measuring device that tests the descriptive theory through its interactions with the real environment. In this language we must also expand the concept of reliability to include the overall predictive value of this description-code or theory-measurement system. I believe it is then reasonable to associate this overall predictive value with what is called the ‘measure of fitness’ in evolutionary theory.

Finally, at the level of nervous activity in the processes of memory and intellectual theory making, we are again searching for more elegant code structures which allow the maximum predictive reliability over the widest domain, but which can be generated from relatively short symbolic descriptions. Perhaps we could even say that the characteristic sign of biological activity at all levels is the existence of efficient and reliable codes. However, at none of these levels can we evade the basic question of how biological systems achieve the unique reliability of their codes through which they have so clearly distinguished themselves from nonliving matter. Even at the level of memory and consciousness it is possible that single enzymes may provide the crucial transmission links or codes from the senses to the internal descriptions in the brain.

1.16 Summary

We have asked once again the historical question: Are the characteristic processes of biological organisms understandable in terms of the basic laws of physics? I have tried to show that in spite of the many classical models of cellular structures and functions there are severe difficulties in accounting for the reliability of hereditary transmission in terms of the elementary laws of physics. I have proposed that the ultimate source of the unique distinction between living and nonliving matter does not rest on idealized classical models of macromolecules, template replication, or metabolic control, but on the *quantitative reliability of molecular codes which can correlate the contents of a quantum mechanical description with its classical phenotypic expression.* To understand such a correlation between quantum descriptions and the corresponding observable classical event requires a quantum theory of measurement applied to elementary molecular hereditary processes. Such a theory presents serious, though I hope not insurmountable, conceptual and formal difficulties for the physicist. However, in spite of the unsolved theoretical questions we can specify certain necessary conditions for individual molecular coding structures. These conditions suggest that the seat of coding or measurement processes in living matter is the individual non-holonomic enzyme catalyst, although it is likely that other structures in the cell serve to increase the reliability of these codes.

Broadly interpreted, the existence of a molecular code of exceptional reliability is essential not only for the origin of life, but also for the development of the individual, the evolutionary process of natural selection, survival of hereditary traits, and even the symbolic coded descriptions that we call intellectual theories. But whatever level of complexity we study, we may expect to find the conformation-dependent tactic catalyst serving as the most elementary hereditary transmission device. For these reasons I believe that describing such reliable hereditary molecular events in terms of quantum mechanics remains the fundamental problem which we must study, not only for theoretical biology, but perhaps also for a firmer epistemological basis for physical theory itself.

Acknowledgments This work is supported by the Office of Naval Research, Contract Nonr 225 (90), and the National Science Foundation, Grant GB 4121. The paper was prepared while the author was a member of the Center for Theoretical Studies, University of Miami, Coral Gables, Florida.

Notes and References

1. See. e.g. E. P. Wigner, *Proc. Natl. Acad. Sci.* 51 (1964) 956.
2. An example of such a statement is found in J. D. Watson, *The Molecular Biology of the Gene* (W. A. Benjamin: New York, 1965) 67: “Until recently, heredity has always seemed the most mysterious of life’s characteristics. The current realization that the structure of DNA already allows us to understand practically all its fundamental features at the molecular level is thus most significant. We see not only that the laws of chemistry are sufficient for understanding protein structure, but also that they are consistent with all known hereditary phenomena.” A much earlier, pre-DNA optimism toward heredity was expressed by Thomas Hunt Morgan in 1919: “That the fundamental aspects of heredity should have turned out to be so extraordinarily simple supports us in the hope that nature may, after all, be entirely approachable.” Quoted from F. H. C. Crick, *On Molecules and Men* (University of Washington Press: Seattle, 1966.)
3. E.g. from D. C. Phillips, *Scientific American* 215 (1966) 90: “. . . as a result of all the work now in progress we can be sure that the activity of Fleming’s lysozyme will soon be fully understood. Best of all, it is clear that methods now exist for uncovering the secrets of enzyme action.”

4. E.g. from J. C. Kendrew, *Scientific American* 216, no. 3 (1967) 142 (reviewing *Phage and the Origins of Molecular Biology,* J. Cairns, G. Stent. and J. Watson, eds.): “up to the present time conventional, normal laws of physics and chemistry have been sufficient, and at least in the opinion of this reviewer the forward horizon is clear of awkward facts that will require new or paranormal laws for their explanation.”
5. See, e.g. V. M. lngram *The Biosynthesis of Macromolecules* (W. A. Benjamin: New York 1965): M. F. Perutz, *Proteins and Nucleic Acids* (Elsevier: Amsterdam, 1962); J. D. Watson, *The Molecular Biology of the Gene* (W. A. Benjamin: New York, 1965).
6. Attempts to justify chance as an explanation of the origin of life are often made, e.g., G. Wald, *Scientific American* (August 1954) 3. For a physicist’s attitude toward chance as an explanation see P. W. Bridgman, S*cience 123* (1954) 16.
7. E. P. Wigner in *The Logic of Personal Knowledge* (Routledge and Kegan Paul: London 1961) p. 231. Also see P. T. Landsberg, *Nature* 203 (1964) 928.
8. N. Bohr, *Atomic Physics and Human Knowledge* (John Wiley and Sons: New York 1958). See esp. pp. 21 and 101.
9. W. M. Elsasser, *Atom and Organism* (Princeton University Press: 1966); also *The Physical Foundations of Biology* (Pergamon Press: New York, 1958).
10. J. M. Burgers, *Experience and Conceptual Activity* (MIT Press: Mass., 1965).
11. See, e.g. R. M. F. Houtappel, H. Van Dam, and E. P. Wigner, *Rev. Mod. Phys.* 37 (1965) 595. For a shorter discussion see E. P. Wigner, *Science* 145 (1964) 995.
12. J. von Neumann in (L. E. Jeffress, ed.) *Cerebral Mechanisms in Behavior* (John Wiley and Sons: New York, 1951) p. 1. A more technical treatment is J. von Neumann in (A. W. Burks, ed.) *The Theory of Self-replicating Automata* (University of Illinois Press: Urbana, 1966).
13. See, e.g. M. Arbib, *Information and Control* 9 (1966) 177; C. V. Lee in *Mathematical Theory of Automata* (Polytechnic Press: Brooklyn) p. 155; J. W. Thatcher, *ibid*. p. 165. J. Myhill in (M. D. Mesarovic, ed.) *Views on General Systems Theory* (John Wiley: New York, 1964) p. 106.
14. See. e.g. J-P. Changeaux, *Scientific American* 212 (1965) 36.
15. A. Sommerfeld, *Mechanics* (Academic Press: New York 1952) p. 80; E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* 4th ed. (Dover: New York, 1944). ch. VIII.
16. W. R. Stahl and H. E. Goheen *J., Theoret. BioI.* 5 (1963) 266; W. R. Stahl ibid. 14 (1967) 187.
17. E. Schrödinger, *What is Life?* (Cambridge University Press, 1944)
18. See. e.g. E. Frieman and R. Goldman, *J. Math. Phys.* 7 (1966) 2153 and references therein.
19. Two classical references on the subject are: W. Heisenberg, *The Physical Principles of Quantum Theory* (Dover: New York, 1930) esp. ch. I V; J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press. 1955) Two recent discussions are: E. P. Wigner, *Am. J. Phys. 31* (1963) 6; A. Daneri, A. Loinger, and G. M. Prosperi, *Nuc. Phys.* 33 (1962) 297. Broader more conceptual discussions are: N. Bohr in *Atomic Physics and Human Knowledge* (John Wiley: New York, 1958) p. 32; P. K. Feyerabend and G. Sussman, each in S. Korner (ed.) *Observation and Interpretation in the Philosophy of Physics* (Dover: New York,1957).
20. This follows the most widely held interpretation as found. e.g., in von Neumann, *loc. cit.*
21. J. Monod, J-P. Changeaux, and F. Jacob, *J. Molec. BioI.* 6 (1963) 306.
22. D. E. Koshland, Jr. *Fed. Am. Soc. Exp. BioI. Proc.* 23 (1964) 719.
23. R.J. Eden, *Proc. Roy. Soc. 205A* (1951) 583.
24. F. London, *Superfluids,* 2nd ed. vol. I (Dover: New York, 1961) p. 8.
25. W.A. Little, *Phys. Rev.* 134 (1964) A 1416.
26. For example, enzyme catalysis has been studied at very high magnetic fields (~220,000 gauss) by B. Rabinovitch, J. E. Maling and M. Weissbluth, *Biophys. J.* (in press). ibid. 7 (1967) 187. No effects were observed; however, owing to the uncertainties in the theory and the fact that critical fields are higher in small superconductors, these results by no means exclude the possibility of superconductive or superfluid properties in enzymes.

27. Although classical approximation may be useful for many types of biological description, we also expect that the problem of the speed and reliability of codes at quantum mechanical dimensions will not be limited to the evolutionary context. In particular, memory and thought in the brain appear to encounter the same type of difficulties with small size, high capacity, and reliability. But in the case of consciousness there is in addition the more obscure problem of the physical basis of self-reference.
28. A list of abiogenic synthesis experiments to 1964 can be found in H. Pattee in *Advances in Enzymology* (F. Nord, ed.) vol. 27 (John Wiley and Sons: New York 1965) p. 381.
29. M. Idelson and E. R. Blout, Polypeptides XV III, *J. Am. Chem. Soc.* 80 (1958) 2387.
30. H. H. Pattee in (A. D. Ketley. ed.) *The Stereochemistry of Macromolecules* (Marcel Dekker: New York, in press).
31. H. Araki and M. M. Yanase, *Phys. Rev. 120* (1960) 622.
32. J. D. Cowan in (N. Wiener and J. P. Schade. eds.) *Progress in Brain Research* vol. 17 (Elsevier: New York,1965) p. 9.

2 How Does a Molecule Become a Message?

2.1 Introduction

The theme of this symposium is "Communication in Development," and, as an outsider to the field of developmental biology, I am going to begin by asking a question: How do we tell when there *is* communication in living systems? Most workers in the field probably do not worry too much about defining the idea of communication since so many concrete, experimental questions about developmental control do not depend on what communication means. But I am interested in the origin of life, and I am convinced that the problem of the origin of life cannot even be formulated without a better understanding of how molecules can function symbolically, that is, as records, codes, and signals. Or as I imply in my title, to understand origins, we need to know how a molecule becomes a message.

More specifically, as a physicist, I want to know how to distinguish *communication* between molecules from the normal physical *interactions* or forces between molecules which we believe account for all their motions. Furthermore, I need to make this distinction *at the simplest possible level,* since it does not answer the origin question to look at highly evolved organisms in which communication processes are reasonably clear and distinct. Therefore I need to know how messages originated.

Most biologists will say that, while this is an interesting question, there are many problems to be solved about "how life works," before we worry about how it all began. I am not going to suggest that most of the "how it works" problems have been solved, but at the same time I do not see that knowing much more about "how it works" in the current style of molecular biology and genetics is likely to lead to an answer to origin problems. Nothing I have learned from molecular biology tells me in terms of basic physical principles why matter should ever come alive or why it should evolve along an entirely different pathway than inanimate matter.

Reprinted from *Communication in Development*, Anton Lang, Ed. *Developmental Biology Supplement* 3, 1–16 (1969). Academic Press, New York and London.

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7,
DOI 10.1007/978-94-007-5161-3_3,

Furthermore, at every hierarchical level of biological organization we are presented with very much the same kind of problem. Every evolutionary innovation amounts to a new level of integrated control. To see how this integrated control works, that is, to see how the physical implementation of this control is accomplished, is not the same as understanding how it came to exist.

The incredible successes of biochemistry in unraveling the genetic code and the basic mechanism of protein synthesis may suggest that we can proceed to the next hierarchical level with assurance that if we pay enough attention to molecular details, then all the data will somehow fall into place. I, for one, am not at all satisfied that this kind of answer even at the level of replication should be promulgated as the "secret of life" or the "reduction of life to ordinary physics and chemistry," although I have no doubt that some of these molecular descriptions are a necessary step toward the answer. I am even less satisfied that developmental programs will be comprehended only by taking more and more molecular data.

Let me make it quite clear at this point that I believe that all the molecules in the living cell obey precisely the laws of normal physics and chemistry (Pattee 1969). We are not trying to understand molecular structure, but language structure in the most elementary sense, and this means understanding not only "how it works," but how it originated. Nor do I agree with Polanyi's (1968) conclusion that the constraints of language and machines are "irreducible"; although I do believe Polanyi has presented this problem—a problem which is too often evaded by molecular biologists—with the maximum clarity. Whatever the case may be, it is not likely that an acceptable resolution of either origin or reduction problems will come about only by taking more data. I believe we need both a theory of the origin of hierarchical organization as well as experiments or demonstrations showing that the hierarchical constraints of a "language" can actually originate from the normal physical constraints that hold molecules together and the laws which govern their motions.

It is essential in discussions of origins to distinguish the sequence of causal events from the sequence of control events. For example, the replicative controls of cells harness the molecules of the environment to produce more cells, and the developmental controls harness the cells to produce the organism; so we can say that development is one level higher than replication in the biological hierarchy. One might argue then that insofar as developmental messages turn off or on selected genes in single cells according to specific interactions with neighboring cells, they can only be a later evolutionary elaboration of the basic rules of self-replication.

However, I believe we must be very cautious in accepting the conclusion of the evolutionary sequence too generally, and especially in extending it to the origin of life. Single, isolated cells clearly exhibit developmental controls in the growth of their structure, so that messages must be generated by interactions of the growing cell with its own structure, so to speak. But since this characteristic structure is certainly a part of the "self" which is being replicated, it becomes unclear how to separate the developmental from the replicative controls. Furthermore, it is one of the most general characteristics of biological evolution that life has increasingly buffered itself from the changes and ambient conditions of the environments. This buffering is accomplished by establishing hierarchical levels of control that grow more and more distinct in their structure and function as evolution progresses. But we must remember that these hierarchical levels always become blurred at their origin. Therefore, when viewing a highly evolved hierarchical organization we must not confuse the existing control chains in the final hierarchical system with the causal chains or evolutionary sequence of their origin.

Our own symbolic languages have many examples of hierarchical structure which do not correspond to a causal order or the sequence in which the structures appeared (e.g., Lenneberg 1967). The evolution of all hierarchical rules is a bootstrap process. The rules do not create a function—they improve an existing function. The functions do not create the rules—they give the rules meaning. For example, stoplights do not account for how people drive—they help people drive more effectively. Nor does traffic create stop lights. Traffic is the reason why stop lights make sense.

Therefore it is reasonable to consider the hypothesis that the first "messages" were expressed not in the highly integrated and precise genetic code that we find today, but in a more global set of geophysical and geochemical constraints, which we could call the primeval "ecosystem language," from which the genetic code condensed in much the same way that our formal rules of syntax and dictionaries condensed from the functional usage of primitive symbols in a complex environment. If this were indeed the case, then it would be more likely that "developmental replication" in the form of external cycles not only preceded autonomous "self-replication," but may have accounted for the form of the genetic code itself.

2.2 Some Properties of Languages and Symbols

The origin of languages and messages is inseparable from the origin of arbitrary rules. It is a general property of languages and symbol systems that their constraints are arbitrary in the sense that the same function can be accomplished by many different physical and logical structures. For example in the case of human language we find many symbol vehicles and alphabets, many dictionaries and syntactical rules, and many styles of writing, all of which function adequately for human communication. The same is true for the machine languages which man has invented to communicate with computers; and as for the physical embodiment of these language structures it is clear, at least in the case of the machine, that the particular physical structures which perform the logic, memory, reading and writing functions are almost incidental and have very little to do with the essential logical constraints of the language system itself.

The arbitrariness in primitive biological languages is less clear. We know that there are many examples of differing organ design with essentially the same function. On the other hand, the universality of the genetic code could be used as an argument against arbitrariness in biological languages. This would be a weak argument at present, however, since the origin of the code is completely unknown. Furthermore, the only experimental evidence, which is meager, indirectly supports the "frozen accident" theory (Crick 1968), which implies that almost any other code would also work.

The “frozen accident” theory also illustrates what I have found to be a principle of hierarchical structures in general, a principle that may be stated as a principle of impotence: Hierarchical organizations obscure their own origins as they evolve. There are several ways to interpret this. We may think of a hierarchical control as a collective constraint or rule imposed on the motion of individual elements of the collection. For such a constraint to appear as a “rule” it must be much simpler than the detailed motions of the elements. The better the hierarchical rule, the more selective it is in measuring particular details of the elements it is constraining. For example, a good stoplight system does not measure all the dynamical details of the traffic, but only the minimum amount of information about the time and direction of cars which, in principle at least, makes the traffic flow as safely and rapidly as practical. This essential simplification or loss of detail is also what obscures the origin of the rule.

This ill-defined property of simplification is common to all language and machine constraints and hierarchical systems in general—that the essential function of the system is “obscured” by too many details of how it works. One well-known example is our spoken language. If while speaking about these problems I were to begin thinking about the details of what I am saying—the syntax of my sentences, my pronunciation, how the symbols will appear on the printed page—I would rapidly lose the function of communication, which was the purpose of all these complex constraints of the language in the first place. In the same way the function of a computer, or for that matter an automobile or a watch, would be lost if to use them we always had to analyze the mechanical details of their components. I would say that the secret of good communication in general lies in knowing what to ignore rather than in finding out in great detail what is going on.

Therefore as a preliminary answer to our first question of how we distinguish communication between molecules from the normal physical interactions, I suggest that one necessary condition for the appearance of a message is that very *complex interactions lead to* a *very simple result.* The nonliving world, at least as viewed by the physicist, often ends up the other way, with the simplest possible problem producing a very complicated result. The more details or degrees of freedom that the physicist considers in his problem the more complex and intricate becomes the solution. This complexity grows so rapidly with the number of particles that the physicist very quickly resorts to a drastic program of relinquishing *all* detailed knowledge, and then talks only about the statistics of very large aggregations of particles. It is only through some “postulate of ignorance” of the dynamical details that these statistical descriptions can be used consistently. Even so, the passage from the dynamical description to the statistical description in physics poses very deep problems which are unavoidably related to the communication of information or messages from the physical system to the observer (Brillouin 1962). If we accept this general idea that communication is in some way a simplification of a complex dynamical process, then we are led by the origin problem to consider what the simplest communication system can be. Only by conceiving of a language in the most elementary terms can we hope to distinguish what is really essential from the “frozen accidents.”

2.3 What Is the Simplest Message?

The biological literature today is full of words like activator, inhibitor, repressor, derepressor, inducer, initiator, and regulator. These general words describe *messengers,* specific examples of which are being discovered every day. I would simplify the messages in all these cases by saying they mean “turn on” or “turn off.” It is difficult to think of a simpler message. But taken by itself, outside the cell or the context of some language, “turn on” is not really a message since it means nothing unless we know from where the signal came and what is turned on as a result of its transmission. It is also clear that the idea of sending and receiving messages involves a definite time sequence and a collection of alternative messages. “Turn on” makes no sense unless it is related by a temporal as well as by a spatial network. On the other hand, one must not be misled by the apparent simplicity of this message. For when such simple messages are concatenated in networks, logicians have shown us that the descriptive potential of such “sequential switching machines” or “automata” are incredibly rich, and that in a formal sense they can duplicate many of the most complex biological activities including many aspects of thought itself. Almost all molecular biological systems operate in this discrete, on-off mode rather than by a continuous modulation type of control. Since many essential input and output variables are continuous, such as concentration gradients and muscle movements, this poses the serious problem, familiar to logicians as well as computer designers, of transcribing discrete variables into continuous variables and vice versa. The transcription process also determines to a large degree the simplicity as well as the reliability of the function.

If the simplest message is to turn something on, then we also need to know the physical origin and limits of the simplest device that will accomplish this operation. Such a device is commonly called a *switch,* and we shall use this term, bearing in mind that it is defined by its function, not by our design of artificial switches that we use to turn on lights or direct trains. The switch is a good example of an element with an exceedingly simple function—it is hard to imagine a simpler function—but with a detailed behavior, expressed in terms of physical equations of motion, which is exceedingly complex. Switches in certain forms, such as ratchets and Maxwell demons, have caused physicists a great deal of difficulty. In a way, this is contrary to our intuition since even a small child can look at a switch or a ratchet and tell us “how it works.” With considerably more effort, using more sophisticated physical and chemical techniques, it may soon be possible to look at allosteric enzyme switches and explain “how they work.”

We must bear in mind, however, that in both cases there are always deeper levels of answers. For example, the physical description “how it works” is possible only if we ignore certain details of the dynamical motion. This is because the switching event which produces a single choice from at least two alternatives is not symmetrical in time and must therefore involve dissipation of energy, that is, loss of detailed information about the motions of the particles in the switch. As a consequence of this dissipation or loss of detail it is physically impossible for a switch to operate with absolute precision. In other words, no matter how well it is designed or how well it is built, all devices operating as switches have a finite probability of being “off” when they should be “on,” and vice versa. This is not to say that some switches are not better than others. In fact the enzyme switches of the cell have such high speed and reliability compared with the artificial switches made by man that it is doubtful if their behavior can be explained quantitatively in terms of classical models. Since no one has yet explained a switch in terms of quantum mechanics, the speed and reliability of enzymes remains a serious problem for the physicist (Pattee 1968). But even though we cannot yet explain molecular switches in terms of fundamental physics, we can proceed here by simply assuming their existence and consider under what conditions a network of switches might be expected to function in the context of a language.

2.4 What Is the Simplest Natural Language?

We come now to the crucial question. An isolated switch in nature, even if we could explain its origin, would have no function in the sense that we commonly use the word. We see here merely the simplest possible instance of what is perhaps the most fundamental problem in biology the question of how large a system one must consider before biological *function* has meaning. Classical biology generally considers the cell to be the minimum unit of life. But if we consider life as distinguished from nonliving matter by its evolutionary behavior in the course of time, then it is clear that the isolated cell is too small a system, since it is only through the communication of cells with the outside environment that natural selection can take place. The same may be said of developmental systems in which collections of cells create messages that control the replication and expression of individual cells.

The problem of the origin of life raises this same question. How large a system must we consider in order to give meaning to the idea of life? Most people who study the origin of life have made the assumption that the hierarchical structure of highly evolved life tells us by its sequence of control which molecules came first on the primeval earth. Thus, it is generally assumed that some form of nonenzymatic, self-replicating nucleic acid first appeared in the sterile ocean and that by random search some kind of meaningful message was eventually spelled out in the sequence of bases, though it is never clear from these descriptions how this lonely “message” would be read. Alternatively, there are some who believe the first important molecules were the enzymes or the switches, which controlled metabolic processes in primitive cell-like units. I find it more reasonable to begin, not with switching mechanisms or meaningless messages, but rather with a primitive communication *network* which could be called the primeval ecosystem. Such a system might consist of primitive geochemical matter cycles in which matter is catalytically shunted through cell-like structures, which occur spontaneously without initial genetic instructions or metabolic control. In my picture, it is the constraints of the primeval ecosystem, which, in effect, generate the language in which the first specific messages can make evolutionary sense. The course of evolution by natural selection will now produce better, more precise, messages as measured in this ecological language; and in this case signals from the outside world would have preceded the autonomous genetic controls which now originate inside the cell.

But these speculations are not my main point. What I want to say is that a *molecule does not become* a *message because of any particular shape* or *structure* or *behavior of the molecule. A molecule becomes* a *message only in the context of* a *larger system of physical constraints which I have called* a *"language"* in analogy to our normal usage of the concept of message. The trouble with this analogy is that our human languages are far too complex and depend too strongly on the structure and evolution of the brain and the whole human organism to clarify the problem. We are explaining the simplest language in terms of the most complex. Anyway, since the origin of language is so mysterious that linguists have practically given up on the problem, we cannot expect any help even from this questionable analogy. What approaches, then, can we find to clarify what we mean by the simplest message or the simplest language?

2.5 The Simplest Artificial Languages

The most valuable and stimulating ideas I have found for studying the origin of language constraints has come from the logicians and mathematicians, who also try to find the simplest possible formal languages which nevertheless can generate an infinitely rich body of theorems. A practical aspect of this problem is to build a computer with the smallest number of switches which can give you answers to the maximum number of problems. This subject is often called "automata theory" or "computability theory," but it has its roots in symbolic logic, which is itself a mathematical language to study all mathematical languages. This is why it is of such interest to mathematicians: all types of mathematics can be developed using this very general language. The basic processes of replication, development, cognitive activity, and even evolution, offer an intriguing challenge to the automata theorist as fundamental conceptual and logical problems, and also to the computer scientist who now has the capability of experimental study of these simulated biological events. There is often a considerable communication gap between the experimental biologist and the mathematician interested in biological functions, and this is most unfortunate, for it is unlikely that any other type of problem requires such a comprehensive approach to achieve solutions.

But let us return to our particular problem of the origin of language structure and messages. What can we learn from studying artificial languages? As I see it, the basic difficulty with computer simulation is that whenever we try to invent a model of an elementary or essential biological function, the program of our model turns out to be unexpectedly complex if it actually accomplishes the defined function in a realistic way. The most instructive examples of this that I know are the models of self-replication. I shall not discuss any of these in detail, but only give the "results."

It is possible to imagine many primitive types of mechanical, chemical, and logical processes which perform some kind of replication (e.g., Penrose 1958; Pattee 1961; Moore 1962). It is also quite obvious that most of these systems have no conceivable evolutionary potential, nor can one easily add on any developmental elaborations without redesigning the whole system or causing its failure.

The first profound model of a self-replicating system that I know was that of the mathematician John von Neumann (1956), who explicitly required of his model that it be capable of evolving a more elaborate model without altering its basic rules. Von Neumann was influenced strongly by the work of Turing (1937), who carried the concept of computation to the simplest extreme in terms of basic operations with symbols, and showed that with these basic rules one can construct a "universal" machine, which could compute any function that any other machine could compute. Von Neumann also made use of the McCulloch and Pitts (1943) models of neuronal switching networks in his thinking about replication, but he extended both these models to include a "construction" process, which was not physically realistic, but which allowed him to describe a "universal self-replicating automaton" which had the potential for evolution and to which developmental programs could be added without changing the basic organization of the automaton.

But what was the significance of such a model? What impressed von Neumann was the final complexity of what started out as the "simplest" self-replicating machine that could evolve. He concluded that there must be a "threshold of complexity" necessary to evolve even greater complexity, but below which order deteriorates. Furthermore, this threshold appeared to be so complex that its spontaneous origin was inconceivable.

Since von Neumann's work on self-replication, there have been further serious logical attempts to simplify or restate the problem (e.g., Arbib 1967a; Thatcher 1963). Automata theory has also been used to describe developmental processes (e.g., Apter and Wolpert 1965; Arbib 1967b); but the basic results are the same. If the program does anything which could be called interesting from a biological point of view, or if it can even be expected to actually work as a program on any real computer, then such programs turn out to be unexpectedly complex with no hint as to how they could have originated spontaneously. For example, one of the simplest models of morphogenesis is the French Flag problem, in which it is required that a sheet of self-replicating cells develop into the pattern of the French Flag. This can be done in several ways (e.g., Wolpert 1968), but the program is not nearly as simple as one might expect from the simplicity of the final pattern it produces.

It is the common feeling among automata theorists, as well as computer programmers, that if one has never produced a working, developmental, replicative, or evolutionary program, then one is in for a discouraging surprise. To help popularize this fact, Michie and Longuet-Higgins (1966) published a short paper called "A Party Game Model of Biological Replication" which will give some idea of the logic to the reader who has had no computer experience. But as computer scientists emphasize, there is no substitute for writing a program and making it work.

Why are all biological functions so difficult to model? Why is it so difficult to imitate something which looks so simple? Indeed, functional simplicity is not easy to achieve, and very often the more stringent the requirements for simplicity of function, the more difficult will be the integration of the dynamical details necessary to carry out the function. While it is relatively easy to imagine *ad hoc* "thought machines" that will perform well-defined functions, the structure of real machines is always evolved through the challenges of the environment to what are initially very poorly defined functions. These challenges usually have more to do with how the machine fails than how it works. In other words, it is the *reliability, stability,* or *persistence* of the function, rather than the abstract concept of the pure function itself, which is the source of structure. We can see this by studying the evolution of any of our manmade machines. Of course in this case man himself defines the general function, but how the structure of the machine finally turns out is not determined by man alone. The history of timepieces is a good example. It is relatively easy to see superficially with each escapement or gear train "how it works," but only by understanding the requirements of precision and stability for survival, as well as the environmental challenges to these requirements in the form of temperature variations, external accelerations, corrosion, and wear, can we begin to understand the particular designs of escapements, gear teeth, and power trains which have survived.

Our understanding of the genetic code and of developmental programs is still at the "how does it work" level, and although we may be able to trace the evolutionary changes, even with molecular detail, we have almost no feeling for which details are crucial and which are incidental to the integrated structure of the organism. The analytical style of molecular biology, which has brought us to this level, first recognizes a highly evolved function and then proceeds to look at the structures in more and more detail until all the parts can be isolated in the test tube, and perhaps reassembled to function again. But if we wish to explain origins or evolutionary innovations, this style may be backward.

If we believe that selective catalysts or "switching molecules" do not make messages by themselves, then we should study not them by themselves, but in switching networks as they might have occurred in a primitive sterile ecosystem. Nor should we try, if we are looking for origins, to design switching networks to perform well defined functions such as universal self-replication or the development of a French Flag morphology, since there is no reason to expect such functions to exist in the beginning. A more realistic approach would be to ask what behavior of more or less random networks of switching catalysts would appear because of its persistence or stability in the face of surrounding disorder. In other words, we should look not for the elements that accomplish well-defined functions, but for the functions that appear spontaneously from collections of well defined elements. How can this be done?

2.6 The Simulation of Origins

The experimental study of the origin of function or any evolutionary innovation is exceptionally difficult because, to observe such innovation naturally, we must let nature take its course. For the crucial innovations we are discussing, like the origin of molecular messages, language constraints, and codes, nature has already taken its course or is going about it too slowly for us to observe. So again we are left with computer simulation of nature, hoping that the underlying dynamics of the origin of hierarchical organization is so fundamental that it can be observed even in a properly designed artificial environment.

The essential condition for the study of natural origins in artificial machines is that we cannot overdefine the function that we hope will originate spontaneously. In other words, we must let the computer take its own course to some degree. A good example of this strategy has been reported by Kauffman (1969). In this example he constructed a "random network" of "random switches" and then observed the behavior. The switches were random in the sense that one of the 2^k Boolean functions of the k inputs to each switch was chosen at random. Once chosen, however, both the switch function and the network structure connecting inputs and outputs of the switches were fixed.

The significant results were that for low connectivity, that is, two or three inputs per switch, the network produced cycles of activity that were both short and stable—short compared to the enormous number of states, and stable in the sense that the network returns to the same cycle even if a switch in that cycle is momentarily off when it should be on, or *vice versa*. Kauffman pictured his network as a very simple model of the genetically controlled enzymatic processes in the single cell; I believe, however, this type of model would more appropriately represent a primeval ecosystem in which initially random sequences in copolymer chains begin to act as selective catalysts for further monomer condensations. With the allowance for the creation of new switching catalysts, we would expect condensation of catalytic sequences produced by the switching cycles, to act very much like a primitive set of language constraints. The copolymer sequences would then represent a "record" of the cycle structure.

In our own group, Conrad (1969) has taken a more realistic view of the physical constraints that are likely to exist on the primitive sterile earth, as well as the competitive interactions and requirements for growth that must exist between replicating organism in a finite, closed matter system. These competitive growth constraints have been programmed into an evolutionary model of a multi-niche ecosystem with organisms represented by genetic strings subject to random mutation and corresponding phenotypic strings which interact with the other organisms. Although this program includes much more structure than the Kauffman program, neither the species nor the environmental niches are initially constrained by the program, but they are left to find their own type of stability and persistence. The population dynamics is determined, not by solving differential equations that can only represent hypothetical laws, but by actually counting the individuals in the course of evolution of the program. Such a program to a large extent finds its own structure in its most stable dynamical configuration, which we can observe in the course of its evolution.

These computer programs illustrate one approach to the study of the origin of the language constraints we have been talking about. They are empirical studies of the natural behavior of switching networks which do not have specific functions designed into them. This is the way biological constraints must have evolved. But even so, you will ask whether these computer simulations are not too far removed from the biological structures, the cells, enzymes, and hormones that are the real objects of our studies.

This is true—the computer is quite different from a cell—but this disadvantage for most studies of "how it works" is also the strength of such simulation for origin studies. The crucial point I want to make is that the collective behavior we are studying in these models is not dependent on exactly how the individual switches work or what they are made of. We are not studying how the switches work, but how the network behaves. Only by this method can we hope to find developmental and evolutionary principles that are common to all types of hierarchical organizations. Only by studies of this type can we hope to separate the essential rules from the frozen accidents in living organisms.

2.7 The Role of Theory in Biology

There has always been a great difference in style between the physical and biological sciences, a difference which is reflected most clearly in their different attitudes toward theory. Stated bluntly, physics is a collection of basic theories, whereas biology is a collection of basic facts. Of course this is not only a difference in style but also a difference in subject matter. The significant facts of life are indeed more numerous than the facts of inanimate matter. But physicists still hope that they can understand the nature of life without having to learn *all* the facts.

Many of us who are not directly engaged in studying developmental biology or in experimenting with particular systems of communication in cells look at the proliferation of experimental data in developmental biology, neurobiology, and ecology and wonder how all this will end. Perhaps some of you who try to keep up with the literature wonder the same thing. Living systems are of course much more complicated than formal languages or present computer programs, since living systems actually construct new molecules on the basis of genetic instruction. But even with a few simple rules and small memories, we know it is possible to write "developmental" programs that lead to incredibly rich and formally unpredictable behavior (e.g., Post 1943). Therefore in the biological sciences it is not altogether reassuring to find that all our data handling facilities, our journals, our symposia, our mail, and even our largest, quickest computers are overburdened with information. The physicist Edward Condon once suggested that the whole scientific endeavor will come to an end because this "data collection" does not converge. Certainly if our knowledge is to be effective in our civilization, we must see to it that our theoretical conceptions are based on the elements of simplicity that we find in all our other integrated biological functions; otherwise our knowledge will not survive.

What we may all hope is that the language constraints at all levels of biological organization are similar to the rules of our formal languages, which are finite and relatively simple even though they are sufficient to generate an infinite number of sentences and meanings. We must remember, at the same time, that the potential variety of programs is indeed infinite, and that we must not consume our experimental talents on this endless variety without careful selection based on hypotheses which must be tested. Of course we shall need more experimental data on specific messenger molecules and how they exercise their developmental controls. But to understand how the molecules became messages, and how they are designed and integrated to perform with such incredible effectiveness, we must also account for the reliability of the controlling molecules as well as the challenges and constraints of the ecosystem which controlled their evolution. This in turn will require a much deeper appreciation of the physics of switches and the logic of networks.

Acknowledgment This work was supported by the National Science Foundation. Grant GB 6932 of the Biological Oceanography Program in the Division of Biological and Medical Sciences.

References

Apter, M. J., & Wolpert, L. (1965). Cybernetics and development. *Journal of Theoretical Biology, 8*, 244.

Arbib, M. A. (1967a). Some comments on self-reproducing automata. In J. F. Hart & S. Takasu (Eds.), *Systems and computer science* (p. 42). Toronto: University of Toronto Press.

Arbib, M. A. (1967b). Automata theory and development: Part 1. *Journal of Theoretical Biology, 14*, 131.

Brillouin, L. (1962). *Science and information theory* (2nd ed., Chapters 20 and 21). New York: Academic Press.

Conrad, M. E. (1969). *Computer experiments on the evolution of co-adaptation in a primitive ecosystem*. Dissertation, Stanford University. [Published as: Conrad, M., & Pattee, H. H. (1970). Evolution experiments with an artificial ecosystem. *Journal of Theoretical Biology, 28*, 393–409.]

Crick, F. H. C. (1968). The origin of the genetic code. *Journal of Molecular Biology, 38*, 367.

Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology, 22*, 437.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.

Mcculloch, W. S., & Pitts, W. (1943). Logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5*, 115. *A*p. 16.

Michie, D., & Longuet-Higgins, C. (1966). A party game model of biological replication. *Nature, 212*, 10.

Moore, E. F. (1962). Machine models of self-reproduction. *Proceedings of Symposia in Applied Mathematics* (vol. 14, p. 17). Rhode Island: Mathematical Problems in the Biological Sciences, American Mathematical Society, Providence.

Pattee, H. H. (1961). On the origin of macromolecular sequences. *Biophysical Journal, 1*, 683.

Pattee, H. H. (1968). The physical basis of coding and reliability in biological evolution. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Vol. 1, p. 67). Edinburgh: Edinburgh University Press.

Pattee, H. H. (1969). Physical problems of heredity and evolution. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Vol. 2, p. 268). Edinburgh: Edinburgh University Press.

Penrose, L. S. (1958). The mechanics of self-reproduction. *Annals of Human Genetics, 23*, 59.

Polanyi, M. (1968). Life's irreducible structure. *Science, 160*, 1308.

Post, E. L. (1943). Formal reductions of the general combinational decision problem. *American Journal of Mathematics, 65*, 197.

Thatcher, J. W. (1963). The construction of the self-describing Turing machine. In *Proceedings of the Symposium on Mathematical Theory of Automata*, vol. 12. Microwave Research Institute Symposia Series, Brooklyn Polytechnic Press.

Turing, A. M. (1937). On computable numbers, with an application to the Entscheidungs problem. *Proceedings of the London Mathematical Society, 2*(42), 230.

von Neumann, J. (1956). The general and logical theory of automata. Reprinted in *The world of mathematics* (p. 2070), J. R. Newman, Ed. New York: Simon & Schuster.

Wolpert, L. (1968). The French flag problem: A contribution to the discussion on pattern development and regulation. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Vol. 1, p. 125). Edinburgh: Edinburgh University Press.

3 Physical Problems of Decision-Making Constraints

Abstract The question is posed as to whether the behavior of living matter gives us any reason to reconsider fundamental physical principles. How is the problem of language likely to influence our concepts of physics? The problems of neuronal activity are felt to be too complex to confront directly with physical principles. We need to understand the physical basis of all symbolic activity on a more fundamental level.

I am looking for problems of neuronal and organismic behavior that demand the attention of the physicist, not because he hopes he can solve the problem of how the brain works, but because this behavior makes him reconsider some fundamental problems of physics. Perhaps this thinking also will result in a better understanding of how the brain works, but that is not the principle stimulus.

What kinds of problem are of this type? First there is what Wigner (1967) calls "the most fundamental question of all"—the mind-body question. Physicists were forced to review this ancient question when they found that it was impossible to formulate quantum theory without considering the process of observation as a classical event (e.g., Bohr 1958; von Neumann 1955). The difficulty arises when we try to find an objective criterion for deciding when an observation has occurred, or equivalently, to decide when we should change from the quantum language to the classical language in describing an observation. Wigner (1967) has argued that since all inanimate matter must in principle be describable in the quantum language, it must be the consciousness of the observer where the switch to classical language becomes unavoidable. Wigner is therefore led to suggest that at the level of the brain

---

Reprinted from *Physical Principles of Neuronal and Organismic Behavior*, M. Conrad and M. Magar, Eds. New York: Gordon and Breach, 1973, pp. 217–223. Presented at the Coral Gables Conference on the Physical Principles of Neuronal and Organismic Behavior, December 18–20, 1970.

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7,
DOI 10.1007/978-94-007-5161-3_4, © Springer Science+Business Media Dordrecht 2012

where consciousness plays a role the equations of quantum mechanics may have to be modified. Similar doubts, or hopes, have been expressed by molecular biologists when they turn to the problems of neurobiology (e.g., Stent 1969; Delbrlick 1970).

3.1 Life Depends on Records

I have taken the point of view that the question of what constitutes an observation in quantum mechanics must arise long before we reach the complexity of the brain. In fact, I propose elsewhere (Pattee 1971a) that the gap between quantum and classical behavior is inherent in the distinction between inanimate and living matter. To state my argument briefly, I would say that living matter is distinguished from non-living matter by its evolution in the course of time, and that this evolution depends on a degree of constraint in a physical system that enables records of past events to control its future behavior. I argue that the very concept of a record is classical in the same sense that a measurement is classical, both depending on dissipative constraints which reduce the number of alternative types of behavior available to the system. The brain, of course, also makes records and uses them to control the body, but before we decide to study the recording process at this level, it is well to remember that the brain is the latest and probably the most intricate set of coordinated constraints resulting from some 3 or 4 billion years of natural selection of large populations of individuals, each controlled by hereditary memories of enormous capacity. Furthermore, the selection process has taken place through interactions with evolving ecosystems whose dynamics are not yet well understood.

3.2 Life Depends on Coordinated Constraints

In addition to the use of records, there is a second universal property of life which I regard as fundamental to our interpretation of physical laws, and that is the coordination of all biological activities by hierarchical controls. Many biologists do not regard the origin of coordinated or functional behavior in matter as a physical problem since they accept the theory of evolution in the form of survival of the fittest as a sufficient explanation. However, this evades the question of the origin of any level of coordinated activity where new functions appear. Specifically it evades the problem of the origin of life, that is, the origin of a minimal set of coordinated constraints which write and read records. This course of recorded evolution has continued to generate level upon level of coordinated, hierarchical constraints from the rules of the genetic code to the rules of the languages of man, and yet we have almost no evidence and hardly any theory of how even one of these control levels arose. For this reason it appears to me that the significant activities of matter and of the mind are separated by level upon level of integrated control hierarchies with the gulfs between each level still hidden by inscrutable mysteries.

If we are to make any progress at all in confronting basic physical principles with the behavior of such hierarchical organisms, then we must begin at the lowest possible level. I have chosen the concept of decision-making to characterize the basic function of a hierarchical control process. I want to consider the simplest examples of decision making in physical terms in order to see what problems arise. Decision-making is, of course, the principle function of the brain, but that does not mean that the essential physics of the brain's function is best studied by looking at such a complex structure.

3.3 Decisions Require Two Levels of Constraint

What is a decision? A decision is a classification of alternatives according to some rules. A decision implies a two-level process in which a number of alternatives generated at the lower level is reduced by some evaluative rules at the upper level. Why must we call this a 'two-level' process? Why is it not possible to describe, all on the same level, a number of alternative events and rules for determining which event occurs? The necessity for levels of description can be seen roughly in the following way. On the one level, the alternatives must be possible, or in some sense physically representable, for if any alternative were totally impossible then deciding against it is a vacuous process. But on another level, in so far as the rules or constraints of decision-making are effective, some of these alternatives actually became impossible, or at least improbable. Now we cannot speak of an event as being both possible and impossible under the same set of laws or rules, and therefore decision-making must occur at a level using different rules than the level where the rules allow alternatives.

But where do the principles of physics enter the decision-making process? Fundamental laws of motion do not include alternatives. That is, physicists divide the world into initial conditions and laws of motion in such a way that the initial conditions give as complete a picture of the state of the system as possible and the laws of motion tell us that the state of the system will change in time in a deterministic way. This leaves no room for alternatives. In the case that the initial conditions are not known we make the further assumption that they have no inherent regularities and hence we may treat them as random (e.g., Wigner 1964). Any other form of behavior of a physical system requires some additional, auxiliary rules which are represented as equations of constraint. Not all forms of constraint allow alternatives, but only remove or freeze-out specified degrees of freedom for all time. However, decision-making constraints must in some sense distinguish alternative behavior of the system as well as indicate a rule for choosing which alternative is actually followed in the course of time. By formally introducing decision-making constraints as certain time-dependent relations between variables, we appear to be adding new "laws of motion" to the system, but since we know they are the result of only local and arbitrary structures we shall refer to such constraints as rules rather than laws. To the extent that such new rules describe the significant behavior of the system we have no more need of the fundamental physical description in terms of the laws of motion. This is why a computer, when it is represented abstractly as a network of ideal rules or noise-free switches, has nothing to do with physical laws. This separation of law and rule is precisely what I want to avoid in my discussion, for I am trying to understand exactly where physical laws enter the rules of decision-making constraints.

3.4 Classical and Quantum Mechanical Decisions

The first problem of physics, then, is to understand what it means to say that constraints classify dynamical systems so as to allow alternative possibilities, and what type of constraints effectively decide which alternative to follow according to a rule. The generation and reduction of alternatives is closely related to the primitive concepts of state description and measurement. In classical mechanics we conceive of the state of a system as one actual case of a set of possible alternative configurations. All the other alternative configurations are in some sense virtual. That is, we can imagine or describe these alternatives, but they exist only as descriptions. By measurement we decide which alternative is “real”. The process of measurement is therefore accomplished by devices which we would call “decision-making” constraints. However, in the classical picture, the essential point is that the decision is made on the alternative *descriptions,* and therefore it has no necessary effect on the “real” situation. For this reason we say that classical measurement and decision-making need not affect the state of the system in any crucial way.

Quantum mechanics forces us to look at decision-making in a more unified way. The state of a system in quantum mechanics is itself made up of a sum of alternatives. This results from the *principle of superposition* which says that an arbitrary linear sum of possible states is also a possible state, and this principle sharply distinguishes the classical from the quantum concept of state-description and measurement. By a measurement process in quantum mechanics we also decide what state the system is actually in, but since the state itself before measurement consisted of alternatives, we have unavoidably altered this state by the measurement. This alteration is known as the *reduction* of the state vector, but the essential point is that it cannot be accounted for by the equations of motion. Therefore, unlike the classical case, all measurements and decision-making processes in quantum mechanics affect the state of the system in a profound and unavoidable way.

Now, as we said, we are looking for the simplest possible decision-making constraints in a physical system, so we might naturally be led to ask whether this primitive device should be considered as a classical or quantum mechanical system. Certainly in one sense it appears simpler to think of a classical decision-making constraint since that is the only kind that man has been able to manufacture and connect up in functional machines, computers and control devices. It is this classical picture which we have extended by analogy to all levels of the nervous system, from the single nerve cell to the brain, although we still do not know how the basic decision-making constraints work or even what they are.

On the other hand, we could argue that any decision-making device which we describe classically is only understood in an approximate way, and furthermore, for smaller and smaller devices the approximation will become less and less valid as the quantum mechanical aspects of its motion become increasingly significant. For example, considering the enzyme molecule as an elementary decision-making constraint, we find that a classical picture of its chemical structure is conceptually useful, but still totally inadequate when it comes to explaining its catalytic power or its specificity in a quantitative way. However, if we try to express the idea of a decision-making constraint in quantum mechanical language we immediately are confronted with the serious difficulties of the measurement problem which we have already mentioned. Let me summarize this situation again with a bit more physical detail.

3.5 The Measurement Problem

In both classical and quantum mechanics the decision process is a two level process. In classical mechanics the lower level requires a dynamical description where the alternatives are represented by different initial conditions and the upper level requires a statistical description of the measuring device. Any decision which decreases two initially equiprobable initial conditions to one (a binary choice) at the dynamical level must be compensated by an increase in the alternatives (entropy) at the statistical level. This well-known trade of entropy for information is at a cost equivalent to a dissipation of approximately kT per binary decision (bit) (e.g., Brillouin 1962). The constraint which accomplishes this decision must allow more alternatives in the initial configuration of the system than is available under the constrained motion of the system. In other words, the constraint results in fewer degrees of freedom of the dynamical motion than are necessary to specify the configuration of the system. This is called non-holonomic constraint (e.g., Sommerfeld 1952; Whittaker 1944).

In quantum mechanics the lower level is the microscopic dynamical (pure state) description, where the alternatives are represented as a superposition of vectors, and the upper level is described as a measurement which reduces the alternatives by a projection transformation. However, there is no simple way to describe any device which actually accomplishes this measurement process (Daneri et al. 1962). At some stage the description must become classical since the final result of the measurement can be expressed only in ordinary language (Bohr 1958). Attempts to impose non-holonomic equations of constraint only serve to emphasize the difficulty in interpreting measurement and decision processes in quantum theory (Eden 1951).

This severe conceptual and formal difficulty in relating the quantum level of behavior to man's ordinary language for describing classical events is the subject of much discussion which we cannot even summarize here (e.g., see Ballantine 1970 and references therein). There is often the implication that our brains "think classically" because they have only interacted directly with the macroscopic classical world. But in molecular biology we extend this classical thinking right down to the single molecule. For example, we speak of the enzyme molecule as recognizing its substrate in the classical sense of deciding whether any molecule it collides with fits the "description" of the substrate as represented by the shape of the substrate binding site. The indication of a positive decision is the catalytic step. But this is not a valid, empirically testable way of looking at enzyme catalysis. We never actually can measure what is going on dynamically in a single enzyme and substrate molecule; we only measure collective, statistical variables such as rates and concentrations. Therefore we really do not know the nature of the decision-making event. We do not even know if any classical model of an enzyme as a non-holonomic constraint will account for the specificity and speed of its decision-making.

Therefore in spite of very helpful classical models of decision-making constraints such as enzymes, there is still a good possibility that the speed and reliability as well as the coordination of decision-making events in living systems depend on quantum mechanical coherence and that it is precisely this dependence which allows the reliable, persistent and intricate evolution of living matter not found in classical or statistical structures.

The general idea of the dependence of life on quantum mechanical properties is not new. Schrödinger (1944) pointed out that it was really the classical laws that were statistical and that any hereditary memory the size of the gene would have to evade the thermodynamic tendency to disorder by persisting in a quantum mechanical stationary state. He also suggested that the macroscopic or classical order in living systems must somehow be a reflection of this quantum mechanical order, but he gave no suggestion as to how the quantum mechanical order was to act as a constraint on the classical order. London (1961) wondered if the unique quantum mechanical long-range order found in superfluids would not provide the possibility for entirely precise motion of biological molecules isolated from the dissipative processes of classical structures, but again he did not suggest how such motion would act as recording or decision-making constraints.

3.6 The Reading and Writing Problem

My own approach to how physical laws are related to life begins with the fact that living matter is controlled by genetic records. The key problem is not the record vehicle itself—we know DNA structure in great detail—but how this structure comes to be *interpreted* or read out as the overriding hierarchical control on the actions of the organism. I do not mean here simply knowing the rules of the genetic code but the actual dynamics of a codon recognition process and the subsequent reaction. The results of the read out process—the actions—we interpret as classical events at all levels, from the choice of a specific amino acid in forming a protein molecule to the brain's choice of words in forming a sentence. By classical events I mean that we do not treat them as superpositions of states but as discrete, definite events which occur with a certain precision or probability. On the other hand, the probability of these events must depend on the detailed dynamics of the read out constraints, and if we are to know the physical basis of life we must find out if these detailed dynamics are consistent with quantum mechanical principles or not.

Wigner (1967) has argued that at the level of consciousness the equations of quantum mechanics may have to be modified by non-linear relations, but since he distinguishes consciousness only in terms of its function as the ultimate observer I see no reason why the same argument should not be used (to the extent that it is valid) at the most elementary level where a classical record results from a quantum mechanical wave function.[1] I would tend to shift this fundamental problem of quantum mechanical measurement away from formalism and look at how any natural record comes to be written and interpreted. What I would like to do is take the physicist, philosopher, and even the biologist out of the picture as far as possible and then ask what meaning we can give to the writing and interpreting of records of the most primitive type. How can we objectively distinguish when a physical structure functions as a record? How do we know when a physical event represents a decision? How can we tell when a molecule contains a message?

3.7 The Language Problem

I cannot imagine any answer to this type of question about the meaning or interpretation of symbolic processes without presupposing some form of generalized language structure. I am thinking now of language in the broadest possible sense, including not only the highly evolved an abstract languages of man and the much more primitive genetic code, but any coordinated set of constraints or rules by which classes of physical structures are transformed into specific actions or events. The essential condition for a language is the *coordination* of its rules, not in the choice of particular rules which generally appear arbitrary. The concept of meaning and interpretation for symbols does not make any sense when applied to single structures, but only to the relations between structures. For this reason I would say that a single decision or record isolated from a set of constraints which can transform classes of such decisions or records into a coordinated activity does not have a meaning or interpretation.

This problem of the *coordination* of decision-making constraints is complementary to the problem of describing the dynamics of a single decision process. To be more explicit, I mean that whether or not a particular dynamical process is a decision or not can only be answered in the context of some coordinated set of language constraints. In exactly the same way, whether or not a particular molecule is a message can only be answered in the context of a coordinated set of rules which forms a language structure (Pattee 1970). For example, whether a DNA molecule is or is not a record or message depends entirely on whether or not there exists a complete set of coding enzymes, RNAs and ribosomes to actually read the message. In the same way, whether or not the detailed dynamics of a gas can be interpreted as a pressure depends on the existence of a constraint which forms a measuring device for pressure. One could object that a real constraint need not exist to define pressure, but only the description of such a constraint. However, the point is that any meaning of description presupposes a language, and this language must at some stage be physically represented as a set of real constraints. In other words, however abstract they may be, all symbols and rules must have a real physical embodiment.

3.8 The Origin Problem

We are led then to our second problem—the physical basis of coordinated constraints which read and write records. Such a coordinated set of constraints I call a generalized language structure. The most universal example of such a language structure is the genetic writing and reading system in which the genetic coding constraints provide the essential read out transformations. What is perhaps most striking about this highly coordinated set of constraints is that it forms the basis for all levels of biological evolution over as long a time span as we can find data, and yet there is no evidence that this set of constraints has itself undergone any significant change. We therefore have theories of biological evolution based on the pre-existence of this genotype-phenotype code, but no idea of how this coordinated set of constraints came into existence (e.g., Crick 1968; Orgel 1968).

This mysterious origin problem is not limited to the genetic code, but is characteristic of all new levels of hierarchical control where a new set of coordinated constraints forms a new language structure which can make decisions about the alternative behavior of the level below. The problem of the origin and nature of coordinated constraints which effectively interpret records and make decisions is therefore a universal problem of all life. Alternatively, we could say that the most fundamental function of coordination in biology is to establish generalized language constraints which allow structure at one level to be interpreted as descriptions and executed as decisions at the higher level. It is in this sense we say that a language system is necessary to establish and execute hierarchical controls (Pattee 1971b).

This association of coordinated constraints with language structures and hierarchical controls does not directly alleviate the problem of origins, but at least it points to the central difficulties. We know some fundamental properties of language structures and control hierarchies which are universal, and we may hope for some hints on how to explain specific origin problems by applying general theories of language and hierarchical control. There are two such universal properties which I believe are particularly significant, the self-referent property and the property of arbitrariness (e.g., Harris 1968).

The self-referent property of language constraints operates at two levels, and for this reason it serves to establish hierarchical control. At the lower level a language structure must be able to describe its own grammar or code rules, and at the upper level it must be able to interpret its own descriptions (i.e., it must contain its own metalanguage). For example, the genetic system not only describes the enzymes and RNAs which execute the coding rules, but also describe control molecules which can interpret messages which turn on and off the expression of structural genes.

Metalanguage statements are necessary if there is to be any coordinated function in addition to self-replication, and hence it is necessary for the evolution of new functions.

A fundamental question is how simple a set of physical constraints can we find which will exhibit these self-referent properties. This in turn requires a clearer understanding of the minimum logical conditions for language. It would also seem to me that any higher cognitive functions must have these self-referent properties in one form or another in order to evolve new levels of hierarchical control, but again the level of complexity of the brain may not be the most efficient place to start (e.g., Minsky 1967).

The property of arbitrariness in symbolic systems is well known but not well understood. Arbitrary is used here in the mathematical sense of chosen precisely, but without further significance to the choice. In physics we might say that initial conditions can be chosen arbitrarily without affecting the equations of motion. In the same way the symbol vehicles of a language may be chosen arbitrarily without affecting the grammar. In biology it is very likely that certain amino acid sequences may be arbitrarily chosen without altering the function of the protein. In the case of the genetic code there is no good evidence either way, but it does not appear unreasonable that life could exist just as well with entirely different codon-amino acid assignments, Similarly we can imagine a reasonably similar form of life with a different set of amino acids or different nucleotides, although at this stage of our knowledge we simply do not know that aspect of biological structure is arbitrary and what is inherently determined.

This property of arbitrariness suggests that the essential properties of coordinated constraints arise only because of the relations or interactions between the constraints which were there from the beginning, and not because of some special structure or property of a particular macromolecule.

One more related observation about the evolution of languages and hierarchical control which may be useful for studying exceedingly complex neuronal organization is that the multiplicity of effective constraints which can perform a type of function decreases as the precision of function increases. In a complementary sense, the multiplicity of effective descriptions of a type of structure decreases as the resolution of the decision-making constraints increases. Consequently, at the earliest stages of new hierarchical controls we may expect a broad class of constraints executing decisions of low specificity. It is important as experimental observers of primitive evolving or learning organizations not to impose our own high resolution classifications to differentiate structures which have no inherent functional difference in the primitive system.

3.9 Summary

To conclude I want to return to the question of whether the behavior of living matter gives us any reason to reconsider fundamental physical principles. How is the problem of language likely to influence our concepts of physics? In the first part of the discussion I tried to show that as long as we used classical dynamical and statistical language we could represent decision-making, recording, and measurement processes as dissipative nonholonomic constraints. However, when we try to describe this type of classification process in the language of quantum mechanics we run into serious conceptual and formal differences. Whatever the formal solution may turn out to be we still end up using classical language to describe the results of a decision, record, or measurement. This suggests that the biological decision-making constraints may have as much to do with our languages of physics as our equations of physics have to do with the language of decision-making, perhaps even more since life arose before physics. In other words, whether or not a particular molecular structure makes a decision may depend on the objective existence of a coordinated set of constraints in the context of which the number of alternative descriptions is actually reduced. Quantum mechanical language would then be a description of matter at a simple enough level that no such constraints can exist. This view does not solve the problem of how coordinated constraints originate, but at least it allows us to view the brain and consciousness as only a very highly evolved and intricate case of such constraints. What we need to do, then, is search for the simplest possible language structures in which decision-making and recording assume an objective meaning, and which hopefully will be simple enough to describe with more clarity than has been possible so far for quantum mechanical measuring processes.

At the present time I feel that problems of neuronal activity are too complex to confront directly with physical principles. We still have too vague a concept of primitive language constraints and too specialized a view of natural laws to recognize the key questions. We need to understand the physical basis of all symbolic activity on a more fundamental level. The history of the matter-mind paradox as well as the measurement problem of quantum mechanics should give us great respect for the difficulties.

Note

1. To my assertion that the incompleteness of quantum mechanical description begins before the phenomenon of consciousness, Wigner replied (private communication): “I believe I understand your arguments in this regard and concur with you. The reason for my arguing on the basis of consciousness was indeed that in this case I could adduce evidence for the incompleteness, whereas I could not do this at a lower level.”

References

Ballantine, L. E. (1970). The statistical interpretation of quantum mechanics. *Reviews of Modern Physics, 42*, 358.

Bohr, N. (1958). *Discussion with Einstein in Atomic Physics and Human Knowledge* (p. 32). New York: Wiley.

Brillouin, L. (1962). *Science and information theory*. New York: Academic.

Crick, F. H. C. (1968). The origin of the genetic code. *Journal of Molecular Biology, 38*, 367.

Daneri, A., Loinger, A., & Prosperi, G. M. (1962). Quantum theory of measurement and ergodicity conditions. *Nuclear Physics, 33*, 297.

Delbrlick, M. (1970). A physicist's renewed look at biology: Twenty years later. *Science, 168*, 1312.

Eden, R. J. (1951). The quantum mechanics of non-holonomic systems. *Proceedings of the Royal Society of London, 205*, 283.

Harris, Z. (1968). *Mathematical structures of language*. New York: Interscience.

London, F. (1961). *Superfluids*. New York: Dover Publications Inc.

Minsky, M. L. (1967). *Computation: Finite and infinite machines* (Chapter 14). Englewood Cliffs: Prentice-Hall Inc.

Orgel, L. E. (1968). Evolution of the genetic apparatus. *Journal of Molecular Biology, 38*, 381.

Pattee, H. H. (1970). How does a molecule become a message? *Developmental Biology, 3*(Suppl.), 1.

Pattee, H. H. (1971a). Can life explain quantum mechanics? In T. Bastin (Ed.), *Quantum theory and beyond* (p. 307). Cambridge: Cambridge University Press.

Pattee, H. H. (1971b). The nature of hierarchical controls in living matter. In R. Rosen (Ed.), *Textbook of mathematical biology* (p. 1). New York: Academic.

Schrodinger, E. (1944). *What is life?* Cambridge: Cambridge University Press.

Sommerfeld, A. (1952). *Mechanics* (Chapter 2). New York: Academic Press Inc.

Stent, G. (1969). That was the molecular biology that was. *Science, 160*, 390.

von Neumann, J. (1955). *The mathematical foundations of quantum mechanics* (p. 417). Princeton: Princeton University Press.

Whittaker, E. T. (1944). *A treatise on the analytical dynamics of particles and rigid bodies* (Chapter 8). New York: Dover Publications.

Wigner, E. P. (1964). Events, laws of nature, and invariance principles. *Science, 145*, 995.

Wigner, E. P. (1967). The problem of measurement. In E. P. Wigner (Ed.), *Symmetries and reflections*. Bloomington: Indiana University Press.

4 Laws and Constraints, Symbols and Languages

In an earlier paper, "How Does a Molecule Become a Message?" written for the 28th Symposium of the Society for Developmental Biology (Pattee 1970a), many of the ideas in my chapter "The Problem of Biological Hierarchy" (Pattee 1970b), in the third volume of *Towards a Theoretical Biology*, were further developed. I would like to add here a summary of my more recent thoughts on these fundamental problems as they have evolved since writing those papers.

In "The Problem of Biological Hierarchy" I state as a central problem of general theoretical biology, "(…) to explain the origin and operation (including the reliability and persistence) of the hierarchical constraints which harness matter to perform coherent functions." I might have better added, the origin, operation, and *evolution* of these constraints, since new hierarchical levels evolve repeatedly. And I conclude there that in order to find an explanation at any level "(…) we will have to understand what we mean by a record or a language (…) ultimately in terms of physical concepts. We will have to learn how collections of matter produce their own internal descriptions."

The difficulty is that any kind of description presupposes some form of language structure. As I conclude in "How Does a Molecule Become a Message?"—"A molecule becomes a message only in the context of a larger system of physical constraints which I have called a "language." Now a language must be a closed set of structures, which we call symbols, some rules for joining and transforming these symbols which we call the grammar, as well as a set of rules for interpreting the meaning of such a collection of these symbols (for example, (Harris 1968)). What has happened is that I have begun with the problem of explaining a *constraint,* which may at first sight appear to be a relatively simple physical concept, but end up trying to explain a *language* which is a very abstract concept that no one fully comprehends. It may seem that I am trying to explain the simple in terms of the complex.

Reprinted from *Towards a Theoretical Biology 4,* C. H. Waddington, Ed. Edinburgh University Press, 1972, pp. 248–258.

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7,
DOI 10.1007/978-94-007-5161-3_5,

I want here, in this present note, to re-emphasize my conviction that dependence on symbol structures and language constraints is the essence of life; that it is in fact the *objective* separation of the simplified symbolic description of the organism from its detailed physical reactions, that is, the separation of genotype and phenotype, which makes the evolution of living systems so profoundly different from non-living systems. Now you will say, what is new about this idea? The genotype and phenotype have been distinguished for well over a hundred years, and it is also well known that evolution depends on this separation. Yes, that much is true, but it is not enough to recognize the universal occurrence of this separation in living matter. To the physicist this separation of genetic description and phenotypic function is not a trivial question, and furthermore the meaning of these concepts does not appear clearer as we discover more details. In fact the acceptance of the structural data of molecular biology as "the physical basis of life" tends to obscure the basic question rather than illuminate it. We are taught more and more to accept the genetic instructions as nothing but ordinary macromolecules, and to forget the integrated constraints that endow what otherwise would indeed be ordinary molecules with their symbolic properties.

What I would like to counteract is the oversimplification, or perhaps what is better called the evasion of the genotype-phenotype distinction. In order to have an explanation of life, this distinction cannot be treated as merely a descriptive convenience for what is popularly assumed to be the underlying molecular basis of life. It is not the structure of molecules themselves, but the internal *self-interpretation* of their structure as symbols that is the basis of life. And this interpretation is not a property of the single molecule, which is only a symbol vehicle, but a consequence of a coherent set of constraints with which they interact.

I would like to explain why the physical concept of constraint especially in the context of origins and evolution, is by no means as simple as elementary physics textbooks lead us to believe. I would also like to suggest though I am far from implementing this suggestion, that the concept of language need not be as complex as linguists teach us. Part of this discrepancy is simply our familiarity with what are usually the most complex language structures, such as man's spoken language, and with very simple constraints, such as a table top. But the concepts of constraint and language are very general, and closely related at a deep level. In fact, I have argued that the most fundamental concept of a constraint in physics depends on an alternative description, and that the apparent simplicity of constraints is in fact a property of the language in which it is described. On the other hand, the most elementary concept of language requires a coherent set of constraints to form its rules of grammar and interpretation.

Thus we have the chicken-egg paradox in a new form: 'Which came first the language or the constraint?'

Laws and constraints. The concept of natural law in physics is quite distinct from the concept of a constraint. A natural law is inexorable and incorporeal, whereas a constraint can be accidental or arbitrary and must have some distinct physical embodiment in the form of a structure. Very often, however, especially in abstract mathematical descriptions of dynamical systems, there is no *formal* way of distinguishing the laws from the constraints. Simply writing down a system's dynamical equations of motion, for example,

$$dx_i / dt = f_i \left(x_i \ldots x_n , t \right)$$

in a general system of coordinates does not reveal whether it is obeying only laws or constraints or both. If we want to predict only the local behavior of the system, this distinction may be unnecessary, but in the context of evolution the only crucial question is how the constraints themselves came about in the course of time.

Constraints, unlike laws of nature, must be the consequence of what we call some form of material structure, such as molecules, membranes, typewriters, or table tops, these structures may be static or time-dependent, but in either case it is important to realize that they are made up of matter which at all times obeys the fundamental laws of nature in addition to behaving as a constraint. What does this mean? If the laws of motion are complete and inexorable, what more can be said? Why isn't an equation of constraint either redundant or inconsistent?

The reason that constraints are not redundant or inconsistent with respect to the laws of motion is that they are *alternative descriptions* of the system. Constraints originate because of a different definition or classification of the system boundaries or system variables even though the equations of constraint may be in the same mathematical form as equations of motion. Our usual justification for choosing to use such auxiliary conditions in place of the detailed dynamics is that it *simplifies* our description of the behavior of the system. For example, the collection of molecules which make up a table top is hopelessly complex as a microscopic dynamical system, but for many practical purposes (that is, its function) it can be alternatively described as a static constraint in the form of a fixed plane, say, $z = \text{constant}$. Fixed constraints are very useful simplifications, but it is the time-dependent constraints, especially the non-integrable constraints, which generate new levels of organization.

Constraints, rules, and regulations. Time-dependent constraints often appear to us as embodiments of rules. Thus a very complex dissipative arrangement of molecules may function as a switch, in which case we often choose to ignore all the dynamical complexity and express the constraint by a simple switching rule. Complex systems may be described by many levels of constraints depending on the degree of simplification which is useful. Thus the abstract switching rule may be useful for a logical analysis of ideal networks, but much too simple for an analysis of the reliability of this function in any real computer.

It is also easy to extend the somewhat subjective concept of *rule* to its more active meaning of *regulation,* which is defined, among other things, as "control according to a rule". Control constraints or mechanisms are, of course, a very complicated and ill-defined set of structures. But in essence control implies that a system possesses *alternative* behaviors, and that owing to the particular nature of the constraint it is possible to correlate a controlling input variable or signal with a particular alternative output dynamics according to a rule.

Again it is important to realize that *controls must operate between different descriptive levels,* just as constraints must be defined by different descriptive levels. This is necessarily the case for all measurement, recording, classification, decision-making, and informational processes in which a number of alternatives on one level of description is reduced by some evaluative procedure at a higher level of description. Why are these necessarily two-level processes? Why are two distinct descriptions necessary? Because we cannot speak of an event as being both possible and impossible using the same level of description. On the lower, unconstrained level the alternatives must be possible; for if they were impossible then deciding for or against them would be a vacuous process. But on the upper, constrained or controlled level, in so far as the rules are reliable or effective, some of these alternatives are actually selected, or more precisely, made more probable, that is, catalyzed (Pattee 1972). This is one fundamental reason for the necessity of hierarchical levels of control which are characteristic of biological organization (Pattee 1970b). I suspect it is also at the root of the measurement problem in which the description of physical events (equations of motion) cannot be used directly for the description of the measurement of these same events (Pattee 1971a). This argument is also very similar to the logician's argument that any description of the truth of a proposition must be in a richer language (metalanguage) than that in which the proposition itself is stated (object language).

But now we are speaking as an outside observer who chooses his descriptions quite subjectively for the purpose of simplifying a problem he wishes to solve or for providing a simple function by a clever use of controls. Living systems on the other hand are created by their constraints and function quite independently of the biologist or physicist who studies them. How do spontaneous constraints arise in matter? Or more exactly, how do we recognize inherent constraints which have evolved autonomously, rather than from the observer's search for simplification or control? How do we distinguish the living system's rules and functions from our subjective attempts to describe such enormously complex systems?

Does life depend on laws or constraints? At this point I believe that the naive realism, characteristic of classical physics and modern biology, runs into serious difficulty when it is applied to explanations or reductions of life to physical laws. Some constraints could in principle be looked at in terms of their detailed dynamics and would be found to obey the laws of motion. This is the way we try to answer the question "How does it work?" when we are presented with a complicated machine or functional constraint. But as I point out in "The Problem of Biological Hierarchy", the question, "How did the functional constraints arise?" cannot be answered so directly because, as we have said, the constraint is not a deterministic consequence of the detailed dynamics of the system but an *alternative description* of the whole system taken as a functional unit.

Now if this alternative description is regarded simply as the outside observer's way of handling the complexity; that is, if the concepts of constraint, function, hierarchical levels, genotype and phenotype, and so on, are regarded as only a useful manner of speaking about the underlying physics and chemistry, then there is no objective difference between living and non-living matter. From this perspective, life appears as nothing but a very complex physical system which we as observers are forced to describe in terms of hierarchical levels of organization. Consequently, in this view the origin of life and its macroscopic evolution is regarded as only a gradual increase in complexity which has necessitated new forms of description on the part of the observing biologist.

This attitude of naive realism runs against almost all modern interpretations of physics. To regard the distinction between genotype and phenotype as based on only a kind of historical biological utility, which can now be replaced more accurately by the 'newly-discovered' underlying physics and chemistry of nucleic acid molecules and enzymes (that is, "the physical basis of life") is very much like claiming that symbols and records can be accurately understood by a detailed physical and chemical analysis of the symbol vehicle or record structure. We know that this is not the case, and that symbol vehicles are largely arbitrary "frozen accidents." It is only the integrated set of rules of grammar and interpretation that gives these particular physical structures their symbolic attributes. What constitutes an "integrated set" or a "language" is of course the basic problem.

In physics, this problem arises in the concept of measurement. Measuring devices are non-integrable constraints which classify and record alternatives, and are not subject to detailed description in terms of the underlying dynamics, even in principle. Measurements are the result of the interaction of a microscopic dynamical system with a special type of dissipative constraint that can so far only be understood by its alternative, statistical description. Furthermore, physicists often regard the dissipative measurement process as more fundamental than the unobservable, formal determinism of the dynamical laws. Thus Born, for example, argues that observation itself is primary: "for whether in a concrete case a cause-effect relation holds can only be judged by applying the laws of statistics to the observations" (Born 1964). And Wigner in a recent review of the epistemology of quantum mechanics concludes: "In my opinion, the restriction of quantum mechanical theory to the determination of the statistical correlations between subsequent observations reproduces most naturally the spirit of that theory" (Wigner 1971).

In any case, whether one regards the laws or the measuring constraints as primary, it is well known among physicists (and should be better known among biologists) that a clear and unified description of events and records of events has not yet proved possible in quantum theory. It is for this reason that constraints in physics need not be so simple; and in fact it is precisely for those nonintegrable constraints which are necessarily associated with the writing and reading of symbols—whether for genetic records, controls, or measurements that unity and objectivity is lacking. Furthermore, if this unity is lacking between events and records of events in physics, then it is not easy to understand how, through the facts of molecular biology, this unification can appear.

On the other hand, by the study of *evolutionary theories*, in particular, theories of the spontaneous origin of life and its hierarchical levels of control, I think we find clues to the solution or at least to the difficulties of finding an objective basis for the separation of matter and symbols. What are some of these clues?

Self-constraint and self-rule. It is my central idea or strategy that *the essence of the matter-symbol problem and hence the measurement or recording problem must appear at the origin of life* where the separation of genotype and phenotype through language structures took place in the most elementary form. Studying the problem in this context helps remove the physicist and logician from the measurement problem. At least it helps widen his tacit anthropocentric assumptions about the function of measurement processes, shifting the emphasis from his brain to the environment or ecosystem where selection takes place. Furthermore, at the level where life originated, the problem must appear greatly reduced in complexity, although it may well be necessary to generalize our concept of records, symbols, and languages to apply them at that primeval level. The relation of this approach to the more conventional physical study of the quantum theory of measurement I have discussed in "Can Life Explain Quantum Mechanics?" (Pattee 1971a).

The necessity of an objective criterion for the occurrence of a separation of events and records has also led me to the hypothesis that *the constraints of living matter must contain their own descriptions.* This follows from the physicist's concept of a constraint as an alternative description of an underlying dynamical process which suffers, as we said earlier, from a basic subjectivity. How do we know that a constraint is nothing more than a convenience for the higher purposes of human computation or control? The only answer I have found is in objectifying the description itself. But of course this shifts the problem to the objective description of the description which at first sight sounds like the beginning of an infinite regress. It is here that we must turn to the fundamental property of a language which, as we said at the beginning, is itself a larger coherent system of constraints. How does this help evade the infinite regress?

If one constraint gets us into this difficulty, how can adding more than one get us out?

This same problem is stated for a language in the following form: How is one symbol or word given a definite meaning? By a coherent set of other words which we call a definition. Then how are the words of the definition given meaning? This also sounds like an infinite regress, but we know that in language this problem is solved. There exist many finite sets of words which can not only define themselves, so to speak, but also define the grammar, as well as form meaningful statements about symbols or groups of symbols, which are called metalinguistic statements. A language therefore possesses the property of self-description or, in the more physical terminology, self-constraint.

There remain, of course, many fundamental questions about language. What is the simplest self-constraining set? Can simpler sets be expected to evolve this self-referent property? How do the grammars of such sets evolve? This type of question tends to emphasize the abstract, symbolic aspects of language, but what I believe more important for biology are the physical properties which make possible symbolic behavior in the first place. Even the most abstract symbols must have a physical embodiment, however arbitrary the symbol vehicle structure may be. Instead of requiring simply a finite, *self-defining* system in the abstract symbolic sense, it is more fundamental to require a finite, *self-constructing* system in the direct physical sense.

This implies a set of constraints which in some co-ordinated way can reconstruct themselves, as well as establish rules by which other structures can be generated. This coordinated set of constraints would amount to a language structure that creates a new hierarchical level of organization by allowing alternative descriptions of the underlying detailed behavior. But the problem of the spontaneous origin of such a set remains.

The origin problem. There are two very general pictures that one may form of how coordinated sets of constraints might arise spontaneously from more or less chaotic beginnings. The most common picture is that of elementary units assembling themselves into larger units. For example, the origin of life is commonly pictured as starting with simple molecules of carbon dioxide, methane, ammonia, and water, and through the activation of some energy source gradually building up amino acids, sugars, bases, then polypeptides and polynucleotides, and finally cell-like aggregations of these macromolecules, until eventually the minimum complexity of living cells is finally reached from which biological evolution can proceed. Some stages of this picture have already been demonstrated in abiogenic experiments. This sequence of origin might be called the formation of *complexity from simplicity.*

A more subtle and less easily pictured process of spontaneous organization can come about the other way around. We imagine initially chaotic aggregations of extreme complexity within which there arise persistent regularities which, so to speak, condense out simple behavior. This formation of *simplicity from complexity* is inherently a collective or global activity of the entire aggregation, as opposed to the locally specific nature of organization created by aggregation of special structures. It is this latter picture of spontaneous organization that accounts more easily for the origin of new hierarchical levels of control, which includes integrated sets of constraints that I have called language structures. For example, while specific enzyme-like polypeptides might arise by the spontaneous-assembly process of our first picture, it is unlikely that this same process could account for the genetic coding enzymes that require an integrated set of highly cooperative constraints.

Unfortunately, specific origins of the second type are much more difficult to model or demonstrate experimentally since they depend on the detailed properties of a globally complex system. On the other hand, the general type of simplification from complexity has been illustrated by models, for example, Thom's catastrophes in his topological dynamics (Thom 1970), dissipative structures in nonequilibrium thermodynamics (for example, (Prigogine et al. 1969)), and Levins' spontaneously simplifying complex systems (Levins 1970). The behavior of Kauffman's randomly connected, random switching nets also illustrates this behavior in fixed sequential systems (Kauffman 1970).

These models and theories are of great help in biology because of their generality and applicability to the problem of the internal generation of constraints. However, they lack many of the known physical and chemical interactions that in my opinion lead to the formation of the most significant constraints. For example, it is hard for me to imagine how the origin of selective, catalytic growth characteristic of present cells can be modeled without including some representation of selective monomer addition steps in individual growing copolymer molecules. In other words, the system must generate internal constraints at the molecular level as well as at the statistical or macroscopic level.

On the other hand, all the demonstrations of abiogenic syntheses at the chemical level, while of great significance for verifying the complexity-from-simplicity theories, lack a realistically complex macroscopic environment. Since the initial conditions of these experiments are kept as simple as possible, they are unlikely to generate constraints of the simplicity-from-complexity type.

Can we test origin theories? How can we design experiments that could demonstrate the spontaneous appearance of constraints that internally simplify the behavior of a complex system? How do we recognize such simple behavior as inherently generated by objective constraints rather than by the outside observer's subjective classifications of the system's behavior?

The first requirement is that the choice of initial conditions be guided by realistic appraisal of the complexity of any primitive earth environment, and not by pre-experimental selection on the part of the designer of what is presumed to be significant. We should think of the initial problem as one of accurately simulating a primitive, sterile ecosystem, rather than biasing our analysis in favor of specific reactions or products. I have for some time suggested this type of simulation as a complement to the many abiogenic synthesis demonstrations (Pattee 1965; Pattee 1971b; Pattee 1973; Pattee 1971c). My first choice would be a simulated sterile seashore with primitive atmosphere, diurnal radiation, sand and clay, waves and tides, and so on, all of which could not have been reasonably missing from the primeval seashore, and all of which seem likely to have significant effects on the chemical development of almost all the reactants, especially any copolymers. Our initial observations of such a system must be more in the style of a naturalist rather than that of a biochemist. But as regularities or predominant behavior appear we must determine if specific catalytic, structural, or thermodynamic constraints are responsible. By their very nature, such simulations are large and expensive and will require cooperative effort from several fields for their design and monitoring. On the other hand, compared to the size and cost of modern exobiology experiments in the form of planetary explorations, such terrestrial simulations are relatively inexpensive and can be expected to provide essential data that could not be obtained any other way.

Even if such realistically complex simulations should develop internal, coordinated constraints relevant to origin problems, this is not equivalent to a theory of origins. A theory need not provide detailed predictions, but at least one might hope that from the theory it would be possible to eliminate unessential complications which may occur in primitive earth simulations. In other words, one would hope to apply the theory to computer programs that might then evolve new hierarchical levels of organization. One difficulty is that all present programming languages operate in a purely symbolic, sequential mode with no necessary physical constraints. As Conrad has suggested in this volume (Conrad 1972), the cost of achieving algorithmic prescription of sequential operations may be the loss of precisely those global condensations that are essential for self-constraint or self-simplification. Whatever the case, we shall learn more about origins only by serious theoretical and experimental study of the problem—the same course that we follow in other sciences.

References

Born, M. (1964). *The natural philosophy of cause and chance* (p. 47). New York: Dover Publications.

Conrad, M. (1972). The importance of molecular hierarchy in information processing. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Essays, Vol. 4, pp. 222–228). Edinburgh: Edinburgh University Press.

Harris, Z. (1968). *Mathematical structures of language*. New York: Wiley.

Kauffman, S. (1970). Behaviour of randomly constructed genetic nets: Binary element nets (p. 18) Continuous element nets (p. 38). In C. H. Waddington (Ed.), *Towards a theoretical biology* (Drafts, Vol. 3). Edinburgh: Edinburgh University Press.

Levins, R. (1970). Complex systems. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Drafts, Vol. 3, p. 73). Edinburgh: Edinburgh University Press.

Pattee, H. (1965). The recognition of hereditary order in primitive chemical systems. In S. Fox (Ed.), *The origin of prebiological systems* (p. 400). New York: Academic Press.

Pattee, H. (1970a). How does a molecule become a message? *Developmental Biology, 3*(Suppl.), 1–16.

Pattee, H. (1970b). The problem of biological hierarchy. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Drafts, Vol. 3, p. 117). Edinburgh: Edinburgh University Press.

Pattee, H. (1971a). Can life explain quantum mechanics? In T. Bastin (Ed.), *Quantum theory and beyond* (p. 307). Cambridge: Cambridge University Press.

Pattee, H. (1971b). The recognition of description and function in chemical reaction networks. In R. Buvet & C. Ponnamperuma (Eds.), *Chemical evolution and the origin of life*. Amsterdam: North Holland.

Pattee, H. (1971c). Physical theories of biological coordination. *Quarterly Reviews of Biophysics, 3*, 255.

Pattee, H. (1972). Physical problems of decision making constraints. *International Journal of Neuroscience, 3*, 99.

Pattee, H. (1973). Physical problems of the origin of natural controls. In A. Locker (Ed.), *Biogenesis evolution and homeostasis*. Heidelberg/New York: Springer.

Prigogine, I., Lefever, R., Goldbetter, A., & Herschkowitz-Kauffman, M. (1969). Symmetry breaking instabilities in biological systems. *Nature, 223*, 913.

Thom, R. (1970). Topological models in biology. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Drafts, Vol. 3, p. 89). Edinburgh: Edinburgh University Press.

Wigner, E. P. (1971). Observations at the end of the conference. Preprint for *On the epistemology of quantum mechanics.* London/Ontario, March 1971.

5 The Physical Basis and Origin of Hierarchical Control

My study of hierarchical organization began from the point of view of a physicist interested in the origin of life. I would not choose to study a concept that is so general and so poorly defined, except that no matter how I have looked at the question of the origin of life, I have not been able to evade the problem of the origin of hierarchical control. Hierarchical organization is so universal in the biological world that we usually pass it off as the natural way to achieve simplicity or efficiency in a large collection of interacting elements. If asked what the fundamental reason is for hierarchical organization, I suspect most people would simply say, "How else would you do it?"

This existential response has the right hierarchical spirit, for indeed one central result of hierarchical organization is greater simplicity; and yet any analytical approach to understanding simplicity always turns out to be very complex. We do not really mean just "simplicity" but *functional* simplicity. The elegance of a physical theory or a work of art depends on simplicity, but never on simplicity alone. There must also be a measure of effectiveness. In the same way, the simplification that results from the hierarchical constraints of an organization must be balanced by how well it functions.

What are the central problems about hierarchical systems? First there is the apparent paradox that hierarchical controls both limit freedom and give more freedom at the same time. The constraints of the genetic code on ordinary chemistry make possible the diversity of living forms. At the next level, the additional constraints of genetic repressors make possible the integrated development of functional organs and multicellular individuals. At the highest levels of control we know that legal constraints are necessary to establish a free society and constraints of spelling and syntax are prerequisites for free expression of thought.

A second problem about hierarchical constraints is that they always appear arbitrary to a large extent. As far as we can see, the same type of life could exist with

---

Reprinted from *Hierarchy Theory: The Challenge of Complex Systems*, H. H. Pattee, Ed. New York: George Braziller, 1973, pp. 73–108.

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7,
DOI 10.1007/978-94-007-5161-3_6,

a number of different genetic codes—that is, with different assignments of nucleic acid codons to amino acids. Molecules that perform the function of messengers, such as hormones or activator molecules, appear to have only arbitrary relation to what they control. Other hierarchical rules are more obviously conventions. We know we can drive on either the left or right side of the road as long as there is collective agreement, just as we know that we can give the same instructions in many different languages using different alphabets. In other words, hierarchical constraints or rules are embodied in structures that are to some extent "frozen accidents."

Even in the arts when there are no chemical or cultural constraints, the artist must invent his own. Igor Stravinsky writes in *Poetics of Music,* "The more constraints one imposes, the more one frees one's self of the chains that shackle the spirit (…)" and he goes on, "(…) the arbitrariness of the constraint serves only to obtain precision of execution."

In the inanimate world, there are certain types of constraints that produce structures that we recognize as atoms, molecules and crystals, and eventually mountains, planets and stars, but these constraints are not associated with new freedom in the collections, and even though some of the shapes of crystals and mountains appear arbitrary, these shapes could not be associated with any kind of effectiveness of function or precision of execution. This is in sharp contrast to all the hierarchical control levels of living matter, and for this reason any theory of hierarchical origins must explain the origin of the type of constraints that are both arbitrary and effective in the sense of giving freedom to the collection.

In contrast to the earlier chapters, it is going to be my strategy to approach these problems primarily from the point of view of physics rather than cellular or developmental biology, sociology, or ecology. At first this may appear as an unlikely strategy since hierarchical organization is foreign to most physics, but common to all the biological sciences. On the other hand, if we want to understand origins we must begin at a simple enough level so that hierarchical controls are not already inherent in the behavior of the system. We must ask what the first level is in the physical world where arbitrary constraints can produce a new kind of freedom of behavior.

5.1 Structural Versus Control Constraints

What I am really asking for is the physical basis for the origin of life, which could be taken as the same problem as the origin of those control constraints that free living matter to evolve along innumerable pathways that non-living matter, following the same detailed laws of motion, cannot follow. In other words, although we recognize structural hierarchies in both living and non-living matter, it is the *control* hierarchy that is the distinguishing characteristic of life. Both structural and control hierarchies are partial orderings, like Simon's Chinese boxes, but the concept of control hierarchy is a narrower partial ordering that implies an active authority relation of the upper level over the elements of the lower levels. It is this original functional meaning I want to preserve, but without the original Divine implications.

Before trying to give a physical description of control, let me briefly review what we mean by a structural hierarchy. Simon, in Chapter I of this volume, has given an excellent description of the physical basis of structural hierarchies as well as a theory of their origin. His theory depends on the relative assembly times for elements with certain probabilities of association and dissociation. This time is drastically shortened if there exist stable substructures. Thus we find development of multicellular organisms based on single cells with high autonomous stability. In the same way, crystals are formed from stable atoms, and words are formed from stable alphabets.

The atom, the molecule, the crystal, and the solid can be distinguished as levels by the criterion of *number*; that is, each level is made up of a large collection of the units of the lower level. However, there is also a physical hierarchy of *forces* underlying these levels, the strongest forces being responsible for the smallest or lowest level structures. The strongest force holds together the nuclei of the atoms, and the weakest force, gravity, holds together the largest bodies of matter. There is also a hierarchy of dynamical *time* scales, which may be associated with the levels of forces, the shortest time with the strongest force and smallest structures, and the longest time with the weakest force and largest structures.

As a result of these graded levels of numbers, forces, and time scales, we can often write dynamical equations for only one level at a time using the approximations that one particle is typical or representative of the collection, that the fast motions one level down are averaged out, and that the slow motions one level up are constant. It is this type of approximate treatment of the dynamics of many hierarchical structures, which Simon calls "near-decomposability." The simplicity and solvability of most physical equations depend on making these approximations. In structural hierarchies this interface can usually be ignored, except for the analysis of the errors of the single-level approximation.

Hierarchical control systems are not this simple. In a control hierarchy the upper level exerts a specific, dynamic constraint on the details of the motion at lower level, so that the fast dynamics of the lower level cannot simply be averaged out. The collection of subunits that forms the upper level in a structural hierarchy now also acts as a constraint on the motions of selected individual subunits. This amounts to a feedback path between levels. Therefore the physical behavior of a control hierarchy must take into account at least two levels at a time, and what is worse, the one-particle approximation fails because the constrained subunits are atypical.

For example, the development of multicellular organisms, which is discussed by Grobstein and Bonner in chapters 2 and 3, shows that the cells do not simply aggregate to form the individual, as atoms aggregate to form crystals. There are chemical messages from the collections of cells that constrain the detailed genetic expression of individual cells that make up the collection. Although each cell began as an autonomous, "typical" unit with its own rules of replication and growth, in the collection each cell finds additional selective rules imposed on it by the collection, which causes its differentiation. Of course, this is also the general nature of social control hierarchies. As isolated individuals we behave in certain patterns, but when we live in a group we find that additional constraints are imposed on us as individuals by some "authority." It may appear that this constraining authority is just one ordinary individual of the group to whom we give a title, such as admiral, president, or policeman, but tracing the origin of this authority reveals that these are more accurately said to be group constraints that are executed by an individual holding an "office" established by a collective hierarchical organization.

In a similar way, developmental controls in cells may be executed by "ordinary" molecules to which we give titles, such as activator, repressor, or hormone, but the control value of these molecules is not an inherent chemical property; it is a complex relation established by a collective hierarchical organization requiring the whole organism. At the lower level of the gene, the authority relation of the hierarchy is often popularly expressed by referring to DNA as the "master molecule" of life, but here again we must emphasize that there is no intrinsic chemical property of DNA that allows it to hold this office. It is the integrated collection of "ordinary" molecules we call the cell that endows DNA with this authority. We should not expect that more detailed study of DNA, enzymes, and hormones would reveal other than ordinary molecules any more than we would expect that a detailed study of presidents would reveal other than ordinary men. The interesting problem of the origin of hierarchical control is to explain how such ordinary molecules and men can evolve such extraordinary authority as members of a collection. Or to put the problem in other words, how do structures that have only common physical properties as individuals achieve special functions in a collection? This statement of the problem shifts the emphasis from one level or another to the hierarchical *interface* between levels.

Rosen (1969) has expressed the problem in a very similar way by defining a hierarchical organization as a system that has more than one simultaneous activity (i.e., structure and function) such that alternative modes of description are an absolute necessity. As I shall show, this is also a characteristic of the physical concept of constraint.

5.2 Structure-Function Problem

For many reasons, I wish I could evade this classical problem of the relation of physical structure to biological function. One reason is that it has generated so many polemics associated with reductionist, vitalist, and teleological arguments. A second reason is that it is currently out of favor as a biological problem. There is some reason for this. Almost all of the discoveries that make up what is called the molecular biological revolution of the last 20 years have been generated by a commitment to a strategy that says that to really understand biological function one must know the molecular details of structure. It was the persistent search for the underlying molecular structures in biochemistry and genetics that has produced our present descriptions of how cells function. Therefore, most biologists today hold strongly to the strategy of looking at the molecular structures for the answers to the question of "how it works."

Nevertheless, it is surprising and discouraging to find so many biologists who, finding this strategy productive, mistake it for a theory of life. Some biology departments have even gone so far as to exclude study of theories of life, as if the detailed facts of molecular biology had somehow demonstrated that theory is not relevant for biology. I was once asked by a leading molecular biologist, quite seriously, "If we can find all the facts, why do we need a theory?" This attitude is especially inappropriate now that molecular biologists are moving on to developmental biology and neurobiology where the integrated function is removed from the detailed structure by even more hierarchical control interfaces. One could not imagine a mathematician trying to understand the nature of computation in terms of how real computer components are designed and wired together. In fact, deep understanding of the nature of computation has come only from theories of computation, which are largely free of the details of real machines. As of now there is no corresponding theory through which we can understand the nature of life, but I shall argue that the study of hierarchical control will form the basis for such a theory.

Let me make it clear that I am not minimizing the importance of collecting the essential detailed biochemical facts of life any more than I would minimize the importance of knowing how to design real switching and memory devices that work with high speed and reliability in a computer. What I wish to emphasize is that the structure-function problem is still very much with us in biology, in spite of our new knowledge of molecular details. I shall try to show that this structure-function duality arises inevitably at a hierarchical control interface. We cannot understand this interface by looking only at structural details or at the functional organization. The problem is precisely at the interface between the detail of structure and the abstraction of function. In fact, what I shall conclude is that function or control can only arise through some selective loss of detail. The problem, which is especially acute for the physicist who believes that nature takes care of all her details, is to explain how a natural "selective loss of detail" can lead to hierarchical control instead of the usual loss of order in the system.

Remember, we are looking for a physical reason why an ordinary molecule can become the controlling factor in forming a chemical bond or in the expression of a whole developmental program. A control molecule is not a typical molecule even though it has a normal structure and follows normal laws. In the collection where it exerts some control it is not just a physical structure—it functions as a *message*, and therefore the significance of this message does not derive from its detailed structure but from the set of hierarchical constraints which we may compare with the integrated rules of a language. These rules do not lie in the structure of any element. We are asking for the physical basis of the hierarchical rules of the collection that turn these ordinary molecules into special messages. Or, to put it in the older terminology, we are still looking for the physical basis of the origin of function in collections of molecules. I prefer, however, to use the concept of hierarchical control rather than hierarchical function, since control implies the constraint or regulation of a total system, whereas function often applies to a specific process that is only a small part of the whole organism. I shall continue, then, by trying to say more clearly what control implies in elementary physical language.

5.3 What Is a Control Device?

How are we to recognize the simplest examples of hierarchical control? How complex must a physical system be in order to exhibit control? Does control appear gradually, or is there a discrete threshold? To answer these questions we must specify more precisely what we mean by "control" in the context of biological origins and evolution. Let me begin to do this with some simple examples. First, consider an idealized process of crystal growth. Suppose you begin with a glass of water with common salt in solution. The sodium and chloride ions are free to move about in three dimensions. We say "free" only in the sense that they each follow the laws of motion of non-interacting particles with only occasional collisions with other molecules or ions. More precisely, we say that most of the time each particle has three translational degrees of freedom. Now suppose that after some time a collection of ions has formed a substantial crystal. This structure can now act as a constraint for some of the ions that land on one of its surfaces. These ions now have fewer degrees of freedom, like a ball constrained to the floor instead of moving through space. It is a collective constraint on individual elements that make up the collection. Do we mean, now, to speak of the crystal as a natural hierarchical control device?

I think not. My purpose in giving this example is to suggest that the concept of control involves a more active dynamical role than simply limiting the available space in which matter can move. The evolution of such time-independent constraints—that is, the addition of more such constraints in the course of time—can lead only to fixed structures, the crystalline solid being a good example. A more realistic variation of this example is the screw-dislocation crystal growth, which is a common process by which crystals actually grow. Instead of each ion binding only at the points of a perfect lattice, there are imperfections in growth which produce a new kind of constraint known as the screw-dislocation. This constraint has two properties: (1) It speeds up the binding of ions by an enormous factor, and (2) it preserves its screw structure as the crystal grows. This persistent speeding up of the rate of growth as the result of the collective structure of the screw dislocation is closer to an active control process.

However, the end result of adding such constraints is still a relatively rigid and permanent structure which clearly does not have the potential for evolution that we associate with living matter. There are many examples of the growth of non-living structures that have this final rigidity. The general difficulty is that we need to find how to add constraints without using up all the degrees of freedom. In other words, what we need for a useful control system is a set of constraints that holds between certain degrees of freedom, but that does not lead to completely rigid bodies. Of course, lack of rigidity is not sufficient; for example, a balloon constrains the gas inside it without freezing up into a rigid body, but a balloon does not have enough constraint on the motions of the gas molecules to be called a control device. A physicist would call the balloon a boundary condition.

What we need to find, then, is a clearer description of the degree of constraint that gives rise to a control hierarchy. We can state two conditions that must be satisfied. First, an effective control event cannot be simply a passive, spatial constraint, but must actively change the *rate* of one particular event, reaction, or trajectory relative to the unconstrained rates. This condition is fulfilled by most devices that we normally associate with existing control systems—for example, switches and catalysts. Second, the operation of the constraint must be *repeatable* without leading to the freezing up of the system. Another way to say this is that control constraints must limit the trajectories of the system in a regular way without a corresponding freezing out of its configurational degrees of freedom. In physical language this condition is satisfied by a *non-holonomic* or non-integrable constraint (e.g., Whittaker 1936). Every interesting man-made control device must be represented as a non-holonomic constraint—switches, ratchets, escapements, and gear shifts being common examples. But at this point we must be very careful not to evade the problem by formalizing it. “Nonholonomic constraint” is acceptable jargon in physics, just as “derepressor” is acceptable jargon in biology. We may state clearly a mathematical function representing a constraint just as we may state clearly the function of a derepressor molecule. But such functions are not derived from what we call the fundamental laws; they are only added on as needed to describe the behavior of integrated systems. In physics the notion of constraint is not considered a fundamental property. It is not useful at the atomic or astronomical scale where the forces between “free” particles are sufficient to describe the motions. So even though we have another word to describe control, we have no idea of how control constraints actually originate. What we need to do is look more closely at the physical basis of natural constraints—how they can arise spontaneously and how they can be classified into structure-producing and control-producing relations.

5.4 What Is a Constraint?

The common language concept of a constraint is a forcible limitation of freedom. This general idea often applies also in mechanics, but as we emphasized in the beginning, control constraints must also create freedom in some sense. Also we must distinguish the forces that enter in the dynamical laws of the system and the forces of constraint. For example, there is the force of gravity and electric fields that enter into the equations of motion and determine how the system will move in the course of time. These fundamental forces do indeed “limit the freedom” of the particles, but the fact is that they leave the particles no freedom at all. Or to put it more precisely, the forces that enter the equations of motion determine the change in time of the state of the system as closely as determinism is allowed by physical theory. The whole concept of physical theory is based on the belief that the motions or states of matter are neither free nor chaotic, but governed by universal laws. So what meaning is there in our concept of additional "constraints" or additional "forceful limitations" on matter when theory says that no additional constraints are possible on the microscopic motions?

The answer is that the physicist's idea of constraint is not a microscopic concept. The forces of constraint to a physicist are unavoidably associated with a new hierarchical level of *description.* Whenever a physicist adds an equation of constraint to the equations of motion, he is really writing in two languages at the same time. The equation of motion language relates the detailed trajectory or state of the system to dynamical time, whereas the constraint language is not about the same type of system at all, but another situation in which *dynamical detail has been purposely ignored*, and in which the equation of motion language would be useless. In other words, forces of constraint are not the detailed forces of individual particles, but forces from collections of particles or in some cases from single units averaged over time. In any case, some form of statistical averaging process has replaced the microscopic details. In physics, then, in order to describe a constraint, one must relinquish dynamical description of detail. A constraint requires an *alternative description.*

Now I do not mean to sound as if this is all clearly understood. On the contrary, even though physicists manage quite well to obtain answers for problems that involve dynamics of single particles constrained by statistical averages of collections of particles, it is fair to say that these two alternative languages, dynamics and statistics, have never been combined in an elegant way, although many profound attempts have been made to do so.[1] Furthermore, the problem has proven exceedingly obscure at the most fundamental level, namely, the interface between quantum dynamics and measurement statistics. This is known as the problem of quantum measurement, and although it has been discussed by the most competent physicists since quantum mechanics was discovered, it is still in an unsatisfactory state. What is agreed, however, is that measurement requires an *alternative description,* which is not derivable from quantum dynamical equations of motion.[2] Bearing in mind that even the clearest physical picture of a constraint involves a hierarchical interface which needs more careful analysis, we may distinguish some of the more common types of structures that we observe to originate spontaneously. We are still looking for conditions that would favor control constraints, but we must begin with the better understood structural constraints.

The chemical bond is undoubtedly the most fundamental structure in the energy range where we recognize life processes. But is it useful to consider this structure as a constraint in the language of control systems? One could argue that since the structures of atoms and molecules are stationary states, they are in a definite state of precise energy, and therefore time is excluded by the uncertainty principle. Thus, we might say that to describe this as a constraint we relinquish the detailed time description of the motion. However, in this case, the structure is really a solution of the equations of motion, and the fact that the energy and time are not simultaneously measurable is not the result of collective forces or an averaging process, but an essential condition of the fundamental dynamical language of quantum mechanics. In other words, the laws determine the structures—there is no alternative. I would, therefore, prefer not to call a single chemical bond a constraint, although it is certainly a structure.

But clearly all constraints depend on chemical bonds. A billiard table is held together by such bonds, and it is called a constraint because the motions of all the atoms in the table can be averaged out to form relatively simple limits on the motion of the billiard balls. So the question arises: How many chemical bonds do we need before we can speak of a constraint? Of course there is no exact number, such as 10 or 10,000. I believe there is only a very pragmatic answer that one can give: I would say that *a dynamical collection is described as a constraint when there exist equations or rules in a simpler form that direct or control the motions of selected particles.* Of course the dynamical equations must still tell us *in principle* how the whole system will evolve in time, without involving the concept of constraint.

In any case, our conclusion is that an equation of constraint in physics in an *alternative description* of the microscopically complex and deterministic motions that gains in simplicity or utility by selectively *ignoring certain dynamical details.* In effect, the physicist has *classified or coded* the microscopic degrees of freedom into a smaller number of new variables. How he performs this classification is not predictable in any dynamical sense, since this depends on the choice of problem he wishes to solve and the degree of approximation he is willing to accept. At the same time, it is also true, as we have mentioned, that the structures of nature fall into "nearly-decomposable" levels that differ in the magnitudes of their forces, sizes, numbers, or time scales, so that the physicist's classifications for a given problem often appear obvious and "natural." But this is a long way from a truly spontaneous classification, which could occur without a highly evolved living system. This kind of classification or selective neglect of detail still requires an intelligent physicist with a problem to solve.

However, I wish to generalize this concept of constraint so that it would apply even before physicists existed. Let me say that a hierarchical constraint is established by a particular kind of new rule that represents not merely a structure but a *classification* of microscopic degrees of freedom of the lower level it controls. The classification may take many forms, which an intelligent observer might define as averages of microscopic variables, or as selection of a few sensitive degrees of freedom, or as a redefinition of the system. But in some sense, the appearance of a natural constraint implies an internal classification process that is selected on the basis of simplicity, utility, or *function* of this alternative description. Now we are ready to ask what types of constraints will lead to hierarchical control. Can we state some principles of hierarchical control?

5.5 The Principle of Classification of Details

By looking at the idea of control and constraint in physical language we have concluded that a hierarchical control level is established by a particular kind of constraint that represents not simply a structure but a classification of the details of the lower level. Before asking what a natural classification could be and how it may arise, I would like to show how this principle of hierarchical control applies generally. Is it reasonable to claim that the selective neglect of certain details is a universal property of hierarchical control systems? Let us consider some examples.

What about the hierarchical control of the developmental processes? Here the lowest level is the collection of self-replicating cells. Bonner found that to represent the developmental process by a program it was necessary to use the concept of the *developmental test.* According to this concept, the developing organism performs tests of the environment or surrounding cells, and the outcome of the tests is to turn off or on the genes appropriate for the developmental response. Now clearly such "tests" must classify interactions. First, there must be a selection of what is tested. For example, such tests would not measure the positions of all the amino acids in the environment—that would hardly be significant for the cell even if it were practical. Second, there must be a selection of what range of results of a test will trigger a control response. Thus, out of the innumerable detailed physical interactions of the cells and their surroundings, there is a classification into significant and insignificant interactions, which I would say amounts to selective neglect of details in favor of only a very limited number of crucial conditions. Goodwin and Cohen (1969) say it is as though every cell need only read a clock and a map, but it is the classification scheme that creates the right "clock" and right "map."

Grobstein's description of the process of folding in the protein molecule provides perhaps the most fundamental example of primitive hierarchical organization. This folding of the linear chain of amino acid residues creates a new property of the chain—the enzymatic function. At the lower level of chain construction, each amino acid is recognized by its transfer enzyme and RNA and placed in sequence under the control of the messenger RNA. Here there is no doubt that certain detailed structures identifying each amino acid must be crucial for accurate construction, whereas on the next level up, when we can use an alternative description of the amino acid collection appropriate to its *function* as an enzyme, such detail is no longer necessary. In fact, there is good experimental evidence that there are equivalence classes of amino acids in the chain which allow the same enzymatic function (e.g., see Nolan and Margoliash 1969).

At the other extreme of complexity, in human organizations and political hierarchies, there are also many examples of the selective ignoring of details. Indeed, at no other level do the rules of classification and selection of what we ignore appear so crucial to the stability and quality of the collective society. It is significant that these human rules about what we ignore are often promulgated as Divine moral principles or at least self-evident conditions for freedom and justice. The principle of "equality under the law" is at the foundation of most legal forms of hierarchical control, and this is little more than an abbreviated statement that legal controls should apply, not to individuals, but to equivalence classes of individuals, where most details must be ignored. An extreme example is the law that says that anyone driving through a red traffic light will be fined. Think of the innumerable details that this control implicitly ignores! Some controls also state explicitly what details we shall ignore, such as laws against racial or religious discrimination.

However, social hierarchies show us that legal controls may either ignore too many details or too few. In the case of traffic control, it is well known that a simple, fixed "constraint" of a four-way stop sign is not as effective in moving traffic as the traffic light that operates according to certain inputs from sensors in the road. In this case, the sensitivity to more details of the motion of vehicles results in increased traffic flow. But it is also clear that a traffic light which took account of too much detail, such as the number of people in the car, could not effectively increase traffic flow. In the same sense, a law that simply requires executing all individuals who have killed another person ignores too many details to be acceptable to society. A few more details must be considered such as the age and mental health of the individual as well as the state of society (i.e., whether the society is "at war"). But again, too much detail would make the control ineffective. The optimum amount of control detail must depend on the desired function. Punishment for murder does not consider the type of weapon used, whereas punishment for robbery does.

5.6 The Principle of Optimum Loss of Detail

We have looked at hierarchical control now at several levels, from crystal growth up through cells to societies. At the extremely simple level of atoms and crystals we argued that the concept of control implied a selective rate control process performed by a collection on individual elements. The control event is repeatable for different elements without freezing up the collection. This type of rate control is accomplished by flexible (non-holonomic) constraints that can be neither too tight nor too loose. If they are too tight, we see more or less rigid bodies with no "function," whereas if they are too loose we see only "boundary conditions" that have no definite, repeatable effect on the elements of the collection.

At the other extreme level of complexity, in human social controls, we also observe that too much constraint on too many details leads to what we might call bureaucratic "freezing up" of the system, whereas too few constraints leads to ineffective function, and ultimately to anarchy. It is easy to imagine how similar difficulties would arise at the enzymatic, cellular, and developmental levels of control if the constraints, or classifications they represent, are either too detailed or too general. Function cannot arise in a strictly deterministic system or in a completely random system.

This situation strongly suggests a second hierarchical control principle. We shall see later that this principle is especially important for understanding hierarchical origins. The principle states that hierarchical control appears in collections of elements within which there is some optimum loss of the effects of detail. Many hierarchical *structures* will arise from the detailed dynamics of the elements, as in the formation of chemical bonds, but the optimum degree of constraint for hierarchical *control* is not determined by the detailed dynamics of the elements. The dynamics of control is determined by how these details are ignored. In other words, *hierarchical controls arise from a degree of internal constraint that forces* *the elements into a collective, simplified behavior that is independent of selected details of the dynamical behavior of its elements.* It is the combination of the *independence* of the constraints on the microscopic dynamics along with the *simplification* of the collective dynamics which creates what we recognize as integrated behavior or function. This optimum level of constraint cannot be specified precisely that is one of our problems—but it falls between the extremes of strict determinism on the one hand and the chaos of "equiprobability of accessible states" on the other. It is significant that the concept of "freedom" has meaning only in behavior that falls between these same extremes, and this helps us understand how freedom can sometimes be generated by adding constraints and other times by removing constraints.

It follows that the origin of an *autonomous* hierarchical control requires that the independence from detail arises spontaneously from this detail. That is, the microscopic dynamics at the lower level must lead to structures with macroscopic dynamical behavior that is independent of microscopic initial conditions, at least over some finite range. What kind of system has this property?

5.7 The Principle of Statistical Closure

When we think of simple dynamical trajectories, such as a baseball thrown to first base or a free satellite coasting toward the moon, we know that success or failure depends crucially on the precision of the initial conditions—that is, the initial position and velocity of the projectile and the position and velocity of its target. This is not the kind of dynamical behavior that leads spontaneously to hierarchical control. At the other extreme, we are familiar with chemical reaction systems in which, no matter what the detailed initial conditions of each reactant molecule, the whole system reaches equilibrium in a very short time. This is not a dynamical but a statistical process, and the theory involved is sometimes characterized as a description of just those aspects of the system that do not depend on initial conditions. Again, we know that this pure statistical behavior does not lead to spontaneous hierarchical control. In dynamical systems, any error in initial conditions eventually leads to chaos, while in statistical systems only chaos in initial conditions leads to simple behavior. What we need for hierarchical control is something in between these extremes, or more precisely, a combination of both.[3]

We have seen that hierarchical controls arise from a collection of elements, but act on individuals of the collection. The spontaneous appearance of collective constraints must therefore be a statistical property of the collection, but its control behavior must act on the dynamics of individual elements. The collective constraint cannot be a dynamical process, since its effect would therefore depend strongly on the initial conditions of its components, which would in effect make it unrecognizable as a constraint. In other words, as we argued earlier, a constraint has no meaning if it cannot be expressed as a simpler alternative description affecting the detailed dynamics. Such an alternative description must have a statistical character, as we have seen, but at the same time this collective behavior of the constraint must harness the detailed dynamical motion of the elements making up the collection.

This harnessing of the lower level by the collective upper level is the essence of hierarchical control. It establishes the "authority" relation between levels within a closed system. It is common to have structural hierarchies beginning with elements and passing to collections of elements, on up through collections of collections, and so on, but such structural hierarchies are open-ended; they do not close back on their lower levels to establish the control-authority relations. (Of course we may also impose controls from outside, but then we are not talking about closed systems or spontaneous origins.) This necessity for the collective, statistical constraint to close back on the internal, detailed dynamics of the elements of the collection I shall call the *Principle of Statistical Closure.*

Now at this point, one might ask if "statistical closure" implies any more than our original definition of a hierarchical control system—namely, a set of collective constraints that in some way limits the detailed behavior of the elements that make up the collection. What I want to do by using this slightly obscure phrase instead of the longer definition is to emphasize what I interpret as the most fundamental aspect of hierarchical control. It is in the nature of a "juxtaposition of disparate categories," to use Koestler's condition for a creative act. Closure is a concept defined in abstract mathematics that is a kind of constraint on the type of rules by which elements of a set can be combined so that no new elements are created. By the use of the closure concept here, I mean a collection of element types that may combine or interact with each other individually in many ways, but that nevertheless persist as the *same* collection when looked at in detail over a long period of time. The mathematical concept of closure is a discrete concept that has nothing to do with the *rate* of combination of the elements in the set. That is, the mathematical idea of closure is entirely independent of time. Similarly, mathematical closure does not involve the possibility of error in local combinations or the number of trials over an extended time interval.

What I mean by *statistical* closure, on the other hand, is a collection of elements that is established and that persists largely because of the *rates* of their combination. This in turn implies a *population dynamics* for the elements and therefore a real-time dependence. Furthermore, the rates of specific combinations of elements must be controlled by collections of the elements in the closed set.

5.8 Examples of Statistical Closure

Since this characterization of hierarchical control is somewhat abstract, I shall give some examples from different levels of biological organization to support the principle that collective statistical constraints do, in fact, establish a closure property on the elements of the hierarchical collection. Perhaps the best known hierarchical control process in biology is at the evolutionary level, where the elements are the individual organisms and the closed collection is the "breeding population." It is a well-known principle of evolution that natural selection does not operate deterministically on individuals, but statistically on the breeding population. The effect of natural selection, however, must be recorded in the hereditary memory of the individual organisms. Therefore, selection is a collective constraint that limits the detailed structure of the individual elements of the collection and establishes a statistically closed set, which we call the breeding population.

At the next lower level we know that the individual is made up of many cells. Here it is not so easy to see what aspects of the development of the multicellular organism are statistical, or even that any statistics are involved at all. Many developmental biologists have adopted the language of computer science to describe the developmental process, as does Bonner in Chapter 3 of this volume, and this might suggest that only deterministic rules are needed to establish the closed collection of cells that forms the individual. Now it may be possible to imagine a strictly discrete, deterministic automaton model of biological development, in which each cell locates its position in the collection of cells by a mechanical test on the fit or state of nearest-neighbor cells, and then turns off or on its appropriate genes according to its developmental program. No statistics would appear to be involved, at least in the formal description of such a model, since the fit or state of other cells might be formally treated as a discrete, mechanical property.

But apparently nature does not do it that way. The individual cell switches off or on or regulates much of its growth according to *concentrations* or *gradients* of concentrations of "message molecules," and this is fundamentally a *statistical* process. Wolpert's (1969) "positional information" is a field concept that requires number averages. Even if some genetic controls are affected by temporal patterns of activity, as in Goodwin's and Cohens's (1969) phase-shift model of development, the concept of *phase* must involve a number or time averaging, which again is a statistical rather than a mechanical process. Since we lack so much detailed knowledge of developmental programs, we cannot go much further with this example. However, from what we know already, the developmental process appears to require statistical constraints formed by a *collection* of cells which establishes a detailed control on the genetic expression of *individual* cells that make up the collection.

Looking now at the cell itself, we reach the most fundamental hierarchical control level, where a small number of monomer types combine to eventually produce a replicating collection. The closure property is obvious, but where are the essential statistical constraints in the cell? Here again, as in the models of development, we can construct self-replicating automata that do not appear to have any statistical process or behavior whatsoever. Even descriptions of self-replication in the language of molecular biology contain no words or concepts that are as obviously statistical as "concentration" or "gradient." Many biology texts presently describe the replication process as entirely deterministic, discrete, mechanical events at the molecular level. Accordingly, one could "design" a cell, using only *one* molecule of each essential nucleic acid and enzyme that would reproduce itself by the same general replication, transcription, translation, and synthesis scheme that is used for real cells, although a certain amount of inefficiency might result. So how can statistical processes arise when only one of each element is necessary?

The answer, I believe, is in the statistical nature of the remarkable molecular devices that make this "scheme" work—the enzymes. Without enzymes there would be no DNA replication, no transcription to RNA, no coding, and no synthesis. The enzymes control the *rate* of selected reactions, and rate control is not a dynamical, but a statistical concept. The reason we do not so easily see the statistical behavior is that selective rate control in enzymes is accomplished by time-averaging rather than number-averaging, which we use to define concentrations. The enzyme (or a more primitive macromolecule with selective catalytic function) is about the simplest, or at least the smallest, structure that fulfills the conditions for a constraint given in a previous section—namely, a complex dynamical system whose functional behavior requires an alternative description of a simpler form. We said this behavior in a hierarchical constraint results in the control of the detailed motions of selected degrees of freedom. This is precisely the nature of enzymes. At the functional level, we speak of the enzyme “recognizing” its substrate, which means that its many collisions with other molecules have no regular effect, and even many degrees of freedom of the substrate are ignored in the binding and catalytic steps. However, at the functional level the operation of the enzyme is specific and repeatable—the same enzyme molecule producing the same catalytic reaction over and over.

Therefore, although any microscopic dynamical description of the enzyme and substrate must follow the reversible laws of motion, at the functional level of description we say the enzyme recognizes the substrate, catalyzes a bond, and resets itself; and all three of these operations are inherently irreversible and hence statistical in some sense. Here both our empirical knowledge of enzymes as well as our ability to construct quantitative models of selective catalysts is weak, and much more thought and experiment will be needed to understand this fundamental type of constraint.

Even though enzyme function is a statistical constraint, it is not by itself an example of statistical closure, since a single enzyme could not selectively constrain all the elements—the amino acids—of which it is made up. In fact, the smallest collection we can now observe which has the statistical closure property is the cell, for in the cell, each monomer of the nucleic acids and proteins is specifically constrained in its reactions by one or more copolymers made from this same collection of monomers. But the cell is too complex to have come into existence spontaneously. *The general hierarchical origin problem* is *to find the simplest or most probable natural collection of elements which exhibits statistical closure.* At the origin of life, the elements were very likely some set of amino acids, nucleotides, or both. At higher levels, of course, the elements are increasingly complex, such as cells or individual organisms. But at each new level there must have been a minimum set of conditions to generate a new statistical closure.

5.9 The Origin Problem

No matter how closely we describe existing hierarchical organizations or how much we appreciate their efficacy, their origins present a new type of problem. First of all, it is difficult to find examples of incipient hierarchical control or observe the spontaneous growth of hierarchical control. Human social hierarchies can be studied historically, but the forces here are so complex and always involve so many intuitive or rational decisions of the human intellect, that any relation with primitive hierarchical origins is thoroughly obscured.

In the field of evolutionary biology we can find many experts who take a rather relaxed and optimistic view of the potential for self-replicating hereditary units to discover new and more effective hierarchical organization through blind mutation and natural selection. On the other hand, there have always been critics who do not find the theory of natural selection a satisfying or sufficient explanation for the *major* evolutionary innovations that have occurred, in particular the origin of new hierarchical control levels. Recently, a new wave of skepticism has appeared among some mathematicians who have tried to imitate or model evolutionary processes on large computers. These attempts include a great variety of problems and programs, which have names such as problem-solving, learning, self-organizing, adaptive, or evolutionary simulations.[4] Even though many of these programs have been extremely sophisticated and have performed well on certain types of life-like problems, their evolutionary potential has been non-existent or disappointing. Of course, these programs have built-in hierarchical structures, but in no case is it obvious that a new level of hierarchical control could ever evolve spontaneously through a blind search and selection process.

The many detailed discoveries of molecular biology have not been much help in clarifying origin problems either. In fact, in one sense, knowing the details has only intensified the problem. Following the ideas of Oparin and Haldane and the successful experiments on the spontaneous synthesis of amino acids and other essential pre-cellular molecules, the origin of life appeared at first as a natural and very likely event. However, now that we know the intricacies of the genetic code and its associated structures, we are again faced with what appears to be an extremely fortuitous hierarchical organization with no obvious process for its spontaneous origin.

Two rather undeveloped theories have been proposed for the universality of the genetic code, and are therefore closely related to its origin. One, called the stereochemical theory, assumes a direct fit between the structure of the three nucleotides forming the codon and the structure of the corresponding amino acid. The other, called the frozen accident theory, assumes that any codon could have been adapted to any amino acid, but that the particular choice was accidental and remains fixed because any change would have been lethal (e.g., Woese 1967; Crick 1968; Orgel 1968).

To my way of thinking, these two theories evade the central problem of the nature of the statistical closure property, which establishes the *coherence* of the code and hence its *descriptive* generality. In the first place, all forms of codes, rules, or descriptions, even the most abstract and symbolic, must have a definite structural basis just as *functions* of all types must have a structural basis. For example, no matter how many adaptor molecules were required between the codon and its amino acid, each adaptor would in some sense possess a structural fit with its neighbor. If this structural fit were only the result of an inherent chemistry of the copolymers involved, then it is by no means clear why any set of copolymers should recognize precisely the set of monomers of which the copolymers are composed. In other words, a stereochemical theory implies that a *structural* closure property is inherent in the physical properties of matter itself. The stereochemical theory is not consistent, then, with the apparent arbitrariness of other types of hierarchical constraints or *descriptions,* or with the idea of *statistical* closure, which depends on the stability or persistence of time-dependent, rate-control processes in populations of elements.

The frozen accident theory, on the other hand, does not conflict with the idea of arbitrariness in the coding rules, but it does not help us understand the statistical closure property either. It is easy to say that a code will be "frozen" when any change is lethal, but the problem arises in the preliminary stages before the coding constraints appear frozen. The establishment of statistical closure may in fact result in an apparent freezing of structures, although there must be earlier levels of freezing before the time the present complex code was established. What we still need to understand are the minimum conditions for the spontaneous appearance of statistical closure in a collection.

It could be that the study of biological organizations in more and more structural detail at all hierarchical levels will eventually suggest the secret of their origins. Certainly such studies are necessary, but I doubt that they will ever be sufficient for a full appreciation of hierarchical origins. The basic reason for my doubt is that detailed structure alone is never a unique basis for any function or description. As we have argued earlier, the concept of hierarchical constraint or function requires an alternative description or a classification of the dynamical details; and the effect or message value of any control molecule is not an inherent property of its structure, but a collective property of a coherent system of molecules. This coherent system allows what we call the alternative *description*, which constrains selected details at the lower structural level so as to establish *function* at the collective upper level. There is, of course, a structural basis for any description or classification rule as well as any functional behavior. When we say that the DNA molecule is a description of the cell or the organism, we imply that there are enzymes, RNAs, and ribosomes to read this description. But it is not the structural details of these molecules that is essential; it is the properties of the entire coordinated set of rate-controlling constraints that make sense of each detail. It is because of the coherence of the set that we can speak of a description, and it is because the meaning of the description is in collective function and not in individual elements that their individual structures remain to some degree arbitrary.

5.10 Summary and Conclusions

The physical basis of hierarchical control does not lie at anyone level of physical theory. This means that we cannot understand the nature of biological hierarchies simply by a finer look at molecular structure, or by solution of detailed equations of motion, nor by application of non-equilibrium statistical thermodynamics. While each of these physical disciplines is useful for describing a particular level of biological organization, hierarchical control operates between levels and is therefore a problem of the nature of the *interface* between levels.

We have characterized this interface by three concepts we call Principles of (1) Classification of Detail, (2) Optimum Constraint, and (3) Statistical Closure.

We assume that at the lower level of any hierarchical organization there is a microscopic dynamics of the elements, described by laws of motion, and deterministic in character. In order to speak of imposing additional "controls" on a system which is already entirely determined in its detailed description, we must relinquish some of this detail and use an alternative description. This amounts to a classification of the microscopic degrees of freedom—our first principle. This classification must be embodied in or executed by what we call a constraint, in particular a flexible (non-holonomic) constraint, which does not simply freeze-out degrees of freedom, but imposes new functional relations between them. The enzyme molecule is our most elementary example of such a flexible constraint that classifies its complex, detailed collisions according to highly simplified functional rules.

We emphasize that the physical description of this type of constraint is necessarily of a *statistical* nature, even though the functional description may appear as a very simple rule. It is this interface between the deterministic laws of the elements and the statistical constraints of their collections for which we still do not have an adequate theory or formalism, and as one might expect, all higher levels of hierarchical control appear to evolve from this primeval level.

Nevertheless, we can recognize several other general properties of hierarchical control systems that hold at all levels. The selection of the classification rules or the degree and type of detail that is ignored in a hierarchical constraint depends on the effectiveness of the particular function that is thereby established. In particular, when the constraints are too numerous or tight, or when they are too scarce or loose, function cannot be maintained. This is our second principle of Optimum Constraint.

Finally, the general property of biological organizations is that they appear to have an indefinite capability to evolve new functions and new hierarchical levels of control while maintaining a relatively fixed set of elementary parts at each level. In other words, the variety of alternative descriptions and functions does not appear to be limited by the fixed set of structural elements that make up the constraints of the organization. We call this Statistical Closure.

A physical theory of the origin of hierarchical control levels would be a derivation of these principles from a combination of the existing fundamental laws, both dynamical and statistical. It would explain how complex collections of interacting elements spontaneously separate out persistent and coherent descriptions and functions under the constraints that relate them. The origin of life is the lowest level of this process where the genotypes (descriptions) and phenotypes (functions) are generated by the constraints of a genetic code. As yet such a physical theory does not exist.

Notes

1. How well the dynamical and statistical descriptions have been combined is, of course, a matter of opinion. The basic problem is that the dynamical equations of matter are strictly reversible in time, whereas collections of matter approaching equilibrium are irreversible. The resolutions of this problem have been central to the development of statistical mechanics and have produced

many profound arguments. For our purposes we need not judge the quality of these arguments, but only note that the resolutions always involve *alternative descriptions* of the same physical situation. (See e.g., Uhlenbeck and Ford 1963)

2. The quantum measurement problem is closely related to the statistical irreversibility problem, and it too has a long history of profound arguments central to the interpretation of quantum theory. The basic difficulty here is that a physical event, such as a collision of two particles, is a reversible process, whereas the record of this event, which we call a measurement, is irreversible (the record cannot precede the event). Yet, if we look at the recording device in detail, it should then be reducible to reversible collisions between collections of particles. Again for our discussion here it is not necessary to judge the many attempts to resolve this difficulty, since as a practical matter they all involve *alternative descriptions* for the event and the record of the event. (See, e.g., von Neumann 1955, for a mathematical treatment, or Wigner 1963, for a non-mathematical review of the problem. For a discussion of quantum measurement and biology, see Pattee 1970).
3. Other types of statistical behavior in between the dynamical and equilibrium statistical systems show a certain degree of self-organization. In strongly nonequilibrium systems, when there is a steady energy flow or dissipation, there is the possibility of oscillations, cycles, and spatial patterns of flow or chemical reaction which could be called hierarchical control (See e.g., Burgers 1963; Morowitz 1968; Prigogine and Nicolis 1971). However, this type of control constrains only statistical variables of the system and does not directly limit the dynamical variables of individual elements of the collection, as in the residue-by-residue ordering in proteins and the switching on or off of single genes in individual cells. (See Pattee 1970, for further discussion of this difference.)
4. Most biological simulations on computers are aimed at understanding or improving our intellectual abilities; e.g., Feigenbaum and Feldman (1963). Several symposia on self-organizing systems, supported by the Office of Naval Research, have covered a wide area of these topics; e.g., Yovits and Cameron (1960), von Foerster and Zopf (1962), Yovits et al. (1962). Examples of more recent computer studies of evolution are Bremermann (1967), Reed et al. (1967), Conrad and Pattee (1970). For a discussion of the theory of evolution between skeptical mathematicians and evolutionary biologists, see Moorhead and Kaplan (1967).

References

Bremermann, H. J. (1967). *Progress in theoretical biology* (Vol. 1). New York: Academic.

Burgers, J. M. (1963). On the emergence of patterns of order. *Bulletin of the American Mathematical Society, 69*, 1.

Conrad, M., & Pattee, H. H. (1970). Evolution experiments with an artificial ecosystem. *Journal of Theoretical Biology, 28*, 393.

Crick, F. H. C. (1968). The origin of the genetic code. *Journal of Molecular Biology, 38*, 367.

Feigenbaum, E. A., & Feldman, J. (Eds.). (1963). *Computers and thought*. New York: McGraw Hill.

Goodwin, B. C., & Cohen, N. H. (1969). A phase-shift model for spatial and temporal organization of developing systems. *Journal of Theoretical Biology, 25*, 49.

Moorhead, P. S., & Kaplan, M. M. (1967). *Mathematical challenges to the NeoDarwinian interpretation of evolution*. Philadelphia: Wistar Institute Press.

Morowitz, H. J. (1968). *Energy flow in biology*. New York: Academic.

Nolan, C., & Margoliash, E. (1969). Comparative aspects of primary structures of proteins. *Annual Review of Biochemistry, 37*, 727.

Orgel, L. E. (1968). Evolution of the genetic apparatus. *Journal of Molecular Biology, 38*, 381.

Pattee, H. H. (1970). Can life explain quantum mechanics? In T. Bastin (Ed.), *Quantum theory and beyond* (p. 307). Cambridge: Cambridge University Press.

Prigogine, I., & Nicolis, G. (1971). Biological order, structure, and instabilities. *Quarterly Reviews of Biophysics, 4*, 107.

Reed, J., Toombs, R., & BarricelIi, N. A. (1967). Simulation of biological evolution and machine learning. *Journal of Theoretical Biology, 17*, 319.

Rosen, R. (1969). Hierarchical organization in automata theoretic models of biological systems. In L. L. Whyte, A. G. Wilson, & D. Wilson, (Eds.), *Hierarchical structures.* New York: American Elsevier, p. 179.

Uhlenbeck, G. E., & Ford, G. W. (1963). *Lectures in statistical mechanics* (Chapter I). Providence, Rhode Island: American Mathematical Society.

von Foerster, H., & Zopf, G. W., Jr. (Eds.). (1962). *Principles of self organization*. New York: Pergamon Press.

von Neumann, J. (1955). *The mathematical foundations of quantum mechanics* (Chapter 5). Princeton: Princeton University Press.

Whittaker, E. T. (1936). *Analytical dynamics* (4th ed., Chapter VIII). New York: Dover.

Wigner, E. (1963). The problem of measurement. *American Journal of Physics, 31*, 6.

Woese, C. R. (1967). *The genetic code*. New York: Harper and Row.

Wolpert, L. (1969). Positional information and the spatial pattern of cellular differentiation. *Journal of Theoretical Biology, 25*(1), 1–47.

Yovits, M. C., & Cameron, S. (Eds.). (1960). *Self-organizing systems*. New York: Pergamon Press.

Yovits, M. C., Jacobi, G. T., & Goldstein, G. D. (Eds.). (1962). *Self-organizing systems 1962*. Washington, DC: Spartan.

6 Postscript: Unsolved Problems and Potential Applications of Hierarchy Theory

Is it possible to have a simple theory of very complex, evolving systems? Can we hope to find common, essential properties of hierarchical organizations that we can usefully apply to the design and management of our growing biological, social, and technological organizations? Such a theory will require a deep and general understanding of the nature of hierarchies, how they originate, how they evolve, how the levels interact, and how failure occurs. The five authors of the previous chapters [Herbert Simon, Clifford Grobstein, James Bonner, Howard Pattee, Richard Levins] have explored the nature of hierarchical organization of complex systems from entirely different perspectives. As mentioned in the Preface, these chapters were developed from a series of public lectures that were scheduled over a period of months, so that the authors did not have the strong interactions that normally occur at conferences. Nevertheless, there is a common theme that runs through all the authors' discussions, and partly because of their independent approaches we may hope this theme has a general significance for a theory of hierarchies.

One purpose of this postscript is to elaborate this theme. Again, because of the different perspectives of the authors, this has required a certain amount of restatement and interpretation which is my own doing and for which the authors cannot be held responsible. A second purpose of this postscript is to suggest directions we may expect the development and applications of hierarchy theory to take. Some of the difficulties are implicit in the previous discussions, but the selection of key problems is also bound to be a matter of individual interpretation. My own selection of unsolved problems is based on the aim of finding new methods for the design and management of very complex systems, since this was the aim that led the sponsors to organize these lectures.

This problem leads to an apparent paradox of any theory of organization. To "manage" any system implies adding at least one hierarchical level—the management level—which oversees the entire system. For this reason, there can never be any closed theory of hierarchies like a dynamical theory in physics. Hierarchy theory

---

Reprinted from *Hierarchy Theory: The Challenge of Complex Systems,* H. H. Pattee, Ed. New York: George Braziller, 1973, pp. 130–156.

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7, DOI 10.1007/978-94-007-5161-3_7, © Springer Science+Business Media Dordrecht 2012

must be more like theories of language or programming that give us useful rules or methods for the most effective design and control of open-ended systems that can continually grow and evolve new levels.

6.1 Common Properties of Hierarchies

The most common and most concrete concept we associate with hierarchical organization is the concept of discrete but interacting levels—what Simon calls "nearly-decomposable" levels. Simon and Grobstein begin with the modified Chinese boxes model of collections within collections, which is an extension of the levels concept to three dimensions or n dimensions. But what generates these levels? Why are the levels discrete? What separates the levels? What couples the levels together? This is the more significant type of question all the authors have tried to approach.

Simon explains or justifies the origin of levels in two ways. First, he argues that the speed of evolution of complex systems will favor those with stable, intermediate *levels of structure.* A difficult construction is easier to accomplish if done in stages. His second argument does not have to do with actually building a complex system; rather, it deals with describing it. Here his point is that complex systems are incomprehensible unless we simplify them by using alternative *levels of description.* His main example is the operation of a computer, which requires several levels of description to program effectively.

Grobstein bases his concept of hierarchy on the developmental process in living organisms. He sees the levels generated by the collective interaction of elementary structures. The transition from the elements to the collection he calls a context change. The structures generated by one set of elements and their interactions become reinterpreted or re-read in the context of a larger set of interactions to form a higher level. Grobstein illustrates this transition by the folding of a linear chain of amino acid residues into a three-dimensional, functioning enzyme, where the new interactions produced by the folding of the chain give a new context to the initial set of amino acids. This folding process is both obvious and mysterious—it is obvious in the structural context, where each interaction is ordinary physics and chemistry, and mysterious in the functional context, which has only been achieved by unknown spans of evolutionary interactions at even higher levels of organization. In any case, Grobstein's basic observation is that the elements of the organization at any level are not limited by their inherent *structural* and interactional properties at that level, but may generate "emergent" properties in the *new context* of a larger set.

Bonner also bases his discussion of hierarchy on the developmental process in living organisms, but he has found it instructive to go beyond the analytical process of isolating the chemicals that control various developmental stages, and ask the synthetic question, "If I were a cell, what control information would I need to produce this organism?" Bonner finds that he cannot imitate the developmental process without levels of alternative subroutines as well as levels of tests for choosing which alternative subroutine to use at a given stage of development. There must be a detailed description within the subroutine level as well as a higher level description of the various subroutines that are “called” according to the results of the developmental tests. In other words, like Simon’s computer program, Bonner finds that the actual construction of the multicellular organism from a single cell requires several *levels of description* in its genetic instructions.

In my own chapter I take up the problem of hierarchical control at the Prebiological level. I ask what physical meaning we can attach to molecules that also serve as messages selectively controlling the rates of alternative events. I find that physical description runs into the same problems as Grobstein’s biological description—namely that molecules have no inherent message, but that message behavior emerges only in the new context of a larger set of constraints. Furthermore, like Simon and Bonner, I find even in physics the transitions to a new context—from the dynamical level to the statistical level, from the statistical level to the constraint level, and from the constraint level to the measurement or control level—all require *alternative levels of description.*

Levins takes up the problem of the limits of complexity at the other extreme of the biological hierarchy—that is, at the ecological and evolutionary level. His basic argument is that very complex systems are dominated by processes that are self-simplifying. While Simon’s point is that complex systems are incomprehensible unless we choose alternative descriptions that selectively ignore detail, Levins argues that complex systems actually tend to persist only in simplified modes of behavior. Levins illustrates this suggestion by the behavior of Kauffman’s randomly connected, random switching networks. These are networks of n switches with two inputs per switch, which have the possibility of passing through 2^n states, but which spontaneously simplify their persistent behavior to approximately $\sqrt{n}$ states—an enormous simplification for large n. Levins further argues that partial decomposability is an *objective* trait of natural selection in complex systems, so that we do not have to view hierarchical organizations as generated from simple components that evolve greater and greater complexity. Instead we may think of hierarchical constraints as *self-simplifications* of initially chaotic, very complex systems (cf. Pattee, 1971). The detailed level of description of the original complex system must then be augmented with an alternative simplified description of the new level.

The common theme characterizing hierarchical systems in these papers that I believe should be emphasized is the double requirement of *levels of description* as well as the more obvious requirement of *levels of structure.* Where there is still a fundamental question—and a wide range of attitudes—is in *the origin and relation of the structural and descriptive levels.* At one extreme there is the view that the laws of nature are entirely objective and determine the structural levels of matter; e.g., particles, atoms, molecules, cells, organisms, etc., which we, as outside observers, find it convenient to analyze by adopting corresponding levels of description. At the other extreme there is the view that whatever the underlying “objective” reality may be, what we understand by laws of nature are structured by the levels of description we choose, and in the context of which we formulate the design and results of our observations.

6.2 The Central Problem of Hierarchy Theory

My own view is that *the relation between the structural and descriptive levels is the central problem* that must be solved to have a theory of hierarchical control. The relation of the structure of the physical world and our observation and description of the world is also the central and profound problem of epistemology, which is normally associated with the philosophy of science rather than the practice of science and its technological applications. One may well ask then, if such questions are important for the very practical aim for which we expect any theory of hierarchical organizations to be useful—that is, the design and management of complex social and technological systems.

At first thought, the epistemological relation between events and descriptions of events—between matter and symbol—might appear largely irrelevant to the problem of programming a computer or managing a factory or an ecosystem. Would we ever seriously evaluate a computer programmer or manager by asking if he is a naive realist or a solipsist? Of course not—at least not in this philosophical context. On the other hand, would we seriously consider hiring a computer programmer who did not have a practical understanding of the difference between hardware and software—that is, between the physical structures of the computer and the descriptive structures by which the programmer controls the computer? We would indeed be upset if we found a programmer who began rewiring the computer circuits the first time his program did not work! On the other hand, he should know that there is always the possibility that the structure of the computer could fail in such a way that no amount of reprogramming could correct the problem. For the same reason we would be skeptical of a new production manager whose first recommendation was to rebuild the factory, although again in rare cases this may be the best alternative. My point is that the manager of any complex organization must clearly appreciate the distinction between the structural levels and the descriptive levels of the system and must know how they interact. He must recognize the difference between a descriptive failure and a structural failure; he must know which levels are dominant and which are subordinate; and he must know when new structural and descriptive levels are needed. Otherwise, even though he is called the manager of the system, he may not have control over it. This we know from experience is often the case with administrators of large social, business, and political organizations.

The primary reason that the modern computer is both a practical machine as well as a useful analogy for other complex organizations is because there is both a clear separation between its structural and descriptive levels—between its hardware and software—as well as a clear understanding of their relationships. The computer hardware is designed with the express purpose of being entirely subordinate to its instructions. One might say that it is one of the purer forms of hierarchical organization, where there is supposed to be absolutely no doubt about where the authority lies or in how the levels are to interact with each other.

However, for all this artificial purity in its hierarchical organization, the largest computers are still no match for natural, biological organizations, especially in those tasks that require creative design, description, and organization. The purity of the artificial levels seems to limit its efficiency. Much of the program of the living organism can now be recognized in the sequence of bases in the DNA molecule of the gene. The sequence of amino acid residues in all the enzymes is known to be subordinate to this sequence of bases, but beyond this, the levels of authority become much more difficult to establish, since the organism lives autonomously as a more or less closed network of control levels. Following Bonner we can say that the expression of the genetic subroutines is controlled by the developmental tests. But by what authority are these tests chosen, described and coordinated? What determines when a new level of structures, descriptions and tests are necessary? What determines the immediate goals or the life-span of the whole organism? What determines the authority levels within a society of organisms? To answer this type of question we need to know how structural and descriptive levels interact. It is not generally the case, in living systems, as it is in the computer, that the descriptive levels dominate the structural levels. The developmental process clearly subordinates the description (i.e., what is expressed) to the organism structure. Similarly the outside environment may at certain times in the life cycle dominate the internal descriptions.

6.3 Extreme Examples of Structure-Description Relationships

The word hierarchy is often applied to ranks or levels of either physical or symbolic structure alone. For example, the sizes of objects from atoms to nebulae can serve as a levels structure (e.g., see Wilson, 1969); or the sizes of symbol structures from letters to paragraphs produce a ranked ordering which can be called hierarchical (e.g., see Whyte, 1969). However, as each author here has made clear in his own way, this somewhat trivial usage of the concept of hierarchy is not of particular interest in the context of a theory of design and control of complex systems. To restate what we said before, to be able to rationally design and control complex organizations will require a deeper understanding of the coupling between physical and symbolic structures—that is, between events and descriptions of events.

It is instructive to look at extreme examples of structural and descriptive interactions. In what type of system do we treat the structure as entirely autonomous and the description as entirely subordinate? Classical concepts of mechanics come very close to this extreme. Space, time, and matter for Newton were absolute structures that had very little to do with their description—so little, in fact, that the interactions of the structure of nature with the description of nature were hardly recognized as a problem. Two centuries later, Hertz clearly recognized the problem in his mechanics, and attempted to clarify it with the requirement of the structure description parallelism; that is, the requirement that the necessary consequences of our descriptions must result in images of nature that are the necessary consequences of the events of nature itself. He also saw that elements of our descriptions could be empty or arbitrary without making the description false, and that the structure of nature did not determine our description of nature unambiguously. Besides parallelism, the only other requirement of the description was that it be perfectly clear and unambiguous in its internal consequences. In any case, we may use the ideal classical picture of nature as an example of structure that was assumed to be largely isolated from its description, and certainly in absolutely no sense influenced or controlled by any description. Dynamical systems theory is essentially an outgrowth of this classical picture of nature.

What example can we find at the other extreme? Are there systems whose structures are entirely subordinate to their descriptions? I would say that an ideal computer would satisfy this requirement. Real computers are designed with this ideal in mind, but obviously cannot achieve the ideal because the structural components are always “limited” in some way by the laws of nature; their size, speed, reliability, for example, are not entirely under descriptive control. Nevertheless, when we program computers or think about the computer as a model or analogue, we cannot at the same time describe its physical structure. That is, in programming we must ignore the physical structure to get logical results. The computer is supposed to do only what we instruct it to do no matter what its structure happens to be. We can imagine, then, at one extreme, the ideal classical “mechanism” that we can describe completely, but over which the description has no influence, and at the other extreme, the ideal computer “mechanism” that we never describe completely but over which we have complete descriptive control.

This contrast can be carried even deeper. We know that the ideal determinism of classical physics is only an approximation that holds for very large objects. In quantum dynamics we find that although we assume a complete mathematical description of the time evolution of pure states that is formally exact and deterministic, it is nevertheless physically impossible to actually measure or control the state of a system with absolute precision. There is the well-known Heisenberg *uncertainty principle,* which expresses the fundamental indeterminism in any attempt to control all the variables defining the state at a given instant of time. On the other hand, we find that purely symbolic or logical systems that are rich enough to allow self-reference (i.e., possessing at least two levels of description) have an essential *incompleteness,* as shown by Gödel. That is, even though the symbols and operations of logic are completely precise and deterministic, there are many types of problems that are unsolvable. For example, there can be no general algorithm for predicting whether a computer will halt, even though the computation is precisely defined.

Therefore, we have the following extreme pictures: On the one hand, there is the very simple quantum mechanical system for which we assume a complete dynamical description in time, but which exhibits a necessary indeterminism when we attempt to observe or control it. On the other hand, we have symbolic logical systems for which we assume precise deterministic operations under our control, but which exhibit incomplete and unpredictable behavior when we try to solve certain types of computational problems. We might say that sufficiently simple natural structures are predictable but uncontrollable, whereas sufficiently complex symbolic descriptions are controllable but unpredictable.

6.4 Thoughts on the Root of the Problem

Real hierarchical systems fall in between these two extremes in the relation between their structural and dynamical levels, but I chose these extreme cases to illustrate the principle difficulty in relating structure and description—or as I have expressed it alternatively, the difficulty in relating *prediction* of how a system behaves and *control* of how it behaves. These pairs of concepts—structure and description, prediction and control are particular examples of hierarchical concepts which have meaning on two different levels. The important point to realize is that all forms of management or control must operate between two hierarchical levels. This is true for any informational constraints where structural alternatives on one level are subordinated by a higher level descriptive process. This is the case in all forms of decision-making, classification, recognition, or measurement process.

Why is this so? Why are two levels of structure and description necessary for any prediction and control process? The basic reason is that in order to predict how a system will behave we must assume it can behave only one way, according to its dynamical law, without the possibility of some alternative behavior. On the other hand, in order to speak of controlling a system we must assume that alternative behaviors are possible. How can a system have control alternatives when no dynamical alternatives exist?

This is the same conceptual problem that has troubled physicists for so long with respect to irreversibility. How can a dynamical system governed deterministically by time-symmetric equations of motion exhibit irreversible behavior? And of course there is the same conceptual difficulty in the old problem of free will: How can we be governed by inexorable natural laws and still choose to do whatever we wish? These questions appear paradoxical only in the context of single-level descriptions. If we assume one dynamical law of motion that is time reversible, then there is no way that elaborating more and more complex systems will produce irreversibility under this single dynamical description. I strongly suspect that this simple fact is at the root of the measurement problem in quantum theory, in which the reversible dynamical laws cannot be used to describe the measurement process. If the event itself is time-symmetric, then the record of the event cannot be, for it is primarily by records that we give time a direction. This argument is also very closely related to the logician's argument that any description of the truth of a symbolic statement must be in a richer metalanguage (i.e., more alternatives) than the language in which the proposition itself is stated.

6.5 More Concrete Examples of the Problem

But now I am straying from the practical questions of the relation of the structural to the descriptive levels. Let us grant that hierarchical control theory must recognize the necessity of alternative levels of description—all the authors have stressed this point.

What has happened to the corresponding structural levels? In my own discussion I have lapsed into the common assumption of a disembodied language or symbol system as a basis for my levels of description. We operate with our own natural language as if it needed no structural basis whatever, and when we speak of other types of symbol systems we usually carry over this unjustified abstraction.

It is because we commonly fail to recognize the nature of the necessary structural basis of symbol vehicles and grammatical constraints that the use of biological organisms and computers as models of hierarchical organizations is so helpful. The perspective of modern biology is entirely structure-oriented—so much so, in fact, that the essential symbolic aspects of life are often missed. DNA is certainly a description, but to most biologists today, DNA is first and foremost a structure. Similarly the genetic code reads and constructs enzymes from this description, but no one imagines these interactions outside the framework of the message and transfer RNA molecules, the coding enzymes, and the ribosomal structures, which actually execute these symbol-manipulating and translating processes. And even when we discuss higher level descriptions—for example, Bonner's developmental subroutines and their tests and transfer commands—there is no doubt that each symbol and operation exists in the cell as a specific molecule or structure. The computer analogy is particularly useful (cf. Simon, p. 12) simply because the levels of structure and levels of description are kept cleanly separated, while at the same time there is no mystery as to what structure corresponds to what symbol or description. However, the computer analogy is not accurate, and this inaccuracy may be a fundamental problem for modeling biological hierarchies. The analogy breaks down because biological structures are grown continuously under the instructions of the genetic program, whereas the structure of computers is relatively fixed. Biological organizations can therefore build new structures from new descriptions, and undoubtedly the richness of the hierarchical levels in living systems depends to some degree on this special ability; but again we have no idea of the processes that generate new levels.

6.6 Theories of Origin of Levels—Instabilities and Catastrophes

How do new levels of hierarchical organization arise spontaneously? How do new structures appear without pre-existing descriptions? This is a fundamental problem for hierarchy theory at all levels. In living organisms it is no exaggeration to say that the genetic code establishes the most fundamental interaction between the descriptive and structural levels, and yet in spite of a rather detailed knowledge of how the code operates, we are still at a loss to understand how the code originated.

At the other end of the evolutionary spectrum, we are now beginning to understand the fundamental levels of language systems, but again we have almost no idea of the origin of these levels. We have learned from our experience with building and managing complex organizations that when the complexity of any level grows beyond a certain range, function becomes impaired, operation becomes inefficient, and reliability declines. We know that *ad hoc* corrections and local improvements in efficiency can only go so far in correcting the problems and that sooner or later we must face a total "reorganization" of the system that must essentially alter the hierarchical control and levels structure. But beyond our traditional empirical knowledge of how such organizations have been run in the past, we are at a loss to design any part of a rational hierarchical structure from theoretical principles.

Levin's paper suggests a theoretical basis for the self-simplification or "clustering" of interactions that would appear as a new hierarchical level; but he assumes an initial complexity that allows natural selection. There are, however, several other theories or approaches to the origin of more elementary physical levels of organization which may turn out to have sufficient depth to qualify for a general theory of hierarchical origins. To conclude this postscript I will try to summarize these approaches and suggest a few possible directions for further development and application.

It is a common observation that the regular dynamical behavior of simple physical systems may suddenly lapse into an entirely new pattern of activity—a spinning top will suddenly fall over, a shade will begin to flap in the wind, or a smooth wave will break into white crests. In dynamical language these sudden changes of pattern we attribute to instabilities. In simple mechanical systems the concept of stability can be clearly defined in terms of the system's ability to return to its original type of trajectory after being transiently perturbed. In a good dynamical theory, the regions of physical instability can be predicted since they correspond to singularities in the mathematical description. But very often the behavior of the system beyond its stable regime is not predictable or even describable in terms of the lower level dynamics. We have an example, then, of a new level of structure and a new level of description arising from instability.

It is important to distinguish this spontaneous generation of a new hierarchical level from the use of new descriptive levels alone. For example, the slow diffusion of dye molecules in a solution can be clearly described in terms of a new concentration variable, but is extremely difficult to describe in terms of the detailed motion of the individual molecules. In this case, however, the motion of the individual molecules is in no way altered or subordinated to the new description using concentration as a variable. In other words, we have levels of description, and if we wish, levels of structure, but no *new* subordination or dominance relations among any of these levels of structure or description. On the other hand, when instabilities occur, the collective motions or patterns act as new constraints on the individual elements and therefore can be said to exercise a dominant effect on their behavior. For example, when we say the flow of molecules in a gas forms a whirlwind, we have more than an alternative description; we have a new structure which dominates the motion of individual molecules. We can say a new hierarchical level has been formed.

The question is how far this process of creation of new levels by instabilities applies to more complex systems. Unfortunately, the very concept of stability loses its clarity as the systems grow in complexity. Furthermore, the theory of dynamical stability is well developed only for local regions of the trajectories, and what any hierarchical theory must recognize is the global stability of the system. Nevertheless, the further study of the behavior of instabilities and singularities in dynamical systems is one of the most promising directions that hierarchy theory should take. This approach has been advanced by Prigogine and his coworkers who have specialized in the description of thermodynamic instabilities far from equilibrium where there is high dissipation and non-linearity. He views the levels of biological hierarchies as closely related to a succession of instabilities (Prigogine and Nicolis, 1971; Glansdorff and Prigogine, 1971).

A more general mathematical approach to the origin of structure through instabilities is Thom's theory of catastrophes, which is based on the concept of structural stability in topological dynamics (Thom, 1970). Thom would not call this a theory of hierarchies, but rather a method for improving our thinking about complex systems—"an art of models." I would agree that we should not expect early advances in our understanding of hierarchical origins to be found in quantitative mathematical models or in new formalisms. We are still too far from a conceptual appreciation—too far from even an intuitive picture of hierarchical complexity. No hierarchy theory will be of much value if it is expressed in a mathematical formalism that is itself of comparable complexity to the systems it is describing. The essence of a good theory of complex systems will tell us how to make them simpler.

6.7 Hierarchy Theory and Systems Theory

How would we expect a theory of hierarchies to differ from the well-established dynamical systems theory and control theory? To what extent should we try to carryover the basic techniques and formalism of systems theory to hierarchies? The dynamical language with its concepts of states, observables, and equations of motion is so general and has proven so useful that it is not likely to be entirely replaced. Nevertheless, dynamical theories are all single-level theories—in one sense a dynamics defines what we call a level. We can use the dynamical language on as many levels as we wish, because we may choose our state variables in almost any way we wish, but for this reason we lose the essential relations between levels. Dynamical theory artificially separates levels by allowing us an unrestricted choice of state variables. The problem of relating the variables on two levels is usually left as a special exercise in interpretation independent of the dynamics at either level. One central aim of control in dynamical systems is to avoid instabilities at all costs, since instability generally leads to a new dynamical regime. A second common goal in control theory is to optimize certain parameters under a given set of fixed constraints. Many of the mathematical techniques of dynamical and control theory have been stimulated by these goals (e.g., see Bellman and Kalaba, 1964).

In terms of structural and descriptive levels, we may look at dynamical formalism as a universal description which, because of its universality, has nothing to do with the particular structure it is describing, other than maintaining the essential Hertzian parallelism; i.e., the consequences of the description must describe the consequences of the natural events. This is the basis of the so-called dynamical analogy which is of great value (e.g., see Rosen, 1970). But this is the same kind of purity we found in the computer program or software, which has nothing to do with the actual structure of the switches and wires of the computer hardware, other than exerting control over the output symbols.

Both dynamics and automata theories, indeed most of mathematics, have been brought to such a state of purity in their descriptive formalisms that hierarchical levels appear totally isolated from each other. Mathematical descriptions therefore tend to create *total* decomposition, whereas the essential behavior of real hierarchical systems depends on the *partial* decomposition of levels, as Simon has emphasized. We find then that dynamical systems theory emphasizes holistic, single-level descriptions, avoidance of instabilities, optimization under fixed constraints and artificial isolation of adjacent levels.

In contrast to systems theory, hierarchy theory must be formulated to describe at least two levels at a time, it must optimize constraints for a given function, and it must allow interactions between alternative levels. Since there is no obvious way to extend the dynamical language to encompass these requirements, perhaps hierarchy theory will require a dualistic or parallel type of theory not unlike the wave-particle duality of quantum physics, where neither description alone is adequate, but where simultaneous use of both appears inconsistent. The essential rules of such a theory would specify under what conditions or for what type of question each description is to be applied.

6.8 Hierarchy Theory and Evolution Theory

Optimization and control must take on a much broader meaning in any theory of hierarchies since all the complex systems we need to control are growing systems. It is not appropriate to consider the constraints as fixed and search for local optima in the parameters. As Levins and I indicated in our chapters, hierarchical control consists largely of optimization of the level of constraints in forming new structural levels and in the optimization of loss of detail in forming new descriptive levels. Also, the theory of hierarchies must be closely related to the theory of evolution. Both must account for the long range behavior of complex, growing systems that have descriptive and structural levels; but while the analogy may be helpful, the theory of evolution itself lacks an independent measure of fitness, without which it has very little predictive value. Therefore, unless we can find better measures of fitness and quality of function we have little hope of achieving any form of optimization in such systems by a theoretical approach. Until we find these measures, tradition and trial and error will continue to be the “method” of hierarchical control as well as evolution.

At this stage in our thinking we are still a long way from a useful theory of complex systems. But our review of the properties of hierarchies offers some suggestions that we may at least use in our explorations or our attempts to control our complex technical, social, or ecological system. First, we should be more cautious in regarding optimization in a systems sense as always a good thing. At least we must be more aware that what is optimized depends on the choice of description. If our description is chosen subjectively or arbitrarily with no necessary integration with the structure it describes, then we have no assurance that the organization as a whole will benefit. An all too common result of attempting to optimize a controllable parameter is to produce an instability that may result in irreversible damage to the organization. On the other hand, a second caution is that not all instabilities should be regarded as a bad thing. While many artificial machines tend to fail if the stable regime of their normal operation is exceeded, natural organizations from cells to ecosystems tend to have many regimes that may be triggered by instabilities. Furthermore, if the theory of organization through instabilities has some truth, then our main effort should be to distinguish those instabilities that are simply disintegrative from those that reintegrate the elements into new levels of organization.

Finally, in hierarchical organizations we should overcome our traditional classical emphasis on the structural levels and recognize the essential role of the descriptive levels in maintaining and coordinating organization. We still put by far our greatest efforts and money into computer hardware, which by now has reached incredible levels of sophistication, rather than into programming theory, which is still in a primitive state. Perhaps for this reason we have not been able to program computers to evolve new levels or model biological evolution in any realistic way. Our preoccupation with the physical and technical aspects of machines has obscured the fact that a failure of description may be more profound and more difficult to correct than a failure of structure. Therefore, along with the study of the hierarchies of biological development, evolution and ecology, physical and mathematical instabilities, and origin processes, the theory of complex systems will also depend on a deeper understanding of descriptive levels and the theory of language.

6.9 Hierarchy Theory and Language Theory

As a final suggestion for directions that the study of hierarchy theories may usefully follow, I would return again to the common problem, the relation between structural and descriptive levels; and now I would ask what meanings we can give to the concept of description at the simplest level. If we want to understand the simplest hierarchical organization we must ask: What is the simplest type of description? The normal usage of the word “description” presupposes a human language in which the description is expressed. However, as we have seen, especially from Bonner’s paper, we also find it meaningful to speak of the cell’s genetic description of itself as well as its description of how to develop into a multicellular organism. Perhaps a linguist would prefer some other word in place of description, such as, prescription or instruction, but that does not touch the root of the problem. Any concept of sign- or symbol-mediated events or behavior presupposes a set of constraints or rules which serve to distinguish the symbols from the events as well as the relation between them. Thus, we call DNA a description only in association with the molecular constraints that make up the genetic code. But in spite of our knowledge of the code and much of its structural embodiment, we have no good theory of its origin or why it has the properties it does. We suspect that much of the structure is arbitrary, just as much of the structural embodiments of human language are arbitrary. Somehow we must find what is essential in both cases. Such an understanding would, in my view, form the foundation of a theory of language as well as the basis for a theory of hierarchies.

I am aware that traditional linguistics would not consider such simple rules as those represented by the genetic code as an example of language. Most linguistic studies tend to define away the problems of origin and evolution of language by refusing to accept non-human symbol systems as language. While this is no place to consider such a difficult matter of definition, it is at least clear that the descriptive potential of the genetic code is rich enough to actually construct in each individual those hierarchical constraints which support language at whatever level one chooses to define the concept. In other words, the descriptive function established by the genetic code, in spite of its apparent simplicity, does not have any obvious limitations in its potential for generating complex hierarchical organizations and functions. Human languages share this apparently unlimited potential for generating complex levels of meanings, but unlike the code, human language is so complicated that as yet we have no clear understanding of the basic elements and rules that are necessary to support these higher languages.

What I suggest therefore, is to study the simplest possible language-like structures or, if you prefer, descriptive processes, that can create a clear separation of hierarchical levels. The genetic code system is much simpler than human language, and yet it is already too complex to understand how it began. Can we imagine simpler rules and relations that allow matter to operate as symbolic representations? Clearly, such simple descriptive systems will not exhibit the range of capabilities of higher language, but at least we should find in such examples a clearer picture of what the physical requirements are for any symbolic activity to arise in matter. Physical theories are created to describe the simplest imaginable systems; for example, the mass point, the hydrogen atom, the equilibrium gas, or the plane wave. The choice of theories also depends strongly on very abstract principles of conservation, symmetry, invariance, or impotence. From these elementary theories we build up descriptions of more and more complex systems. But in all these efforts we take for granted that we may use any language we wish and as many as necessary. That is, we choose whatever mathematical formalism is most useful and then interpret the symbols and measurement operations in very highly developed natural language. To a large degree, the simplicity of natural laws arises through the complexities of the languages we use for their expression (cf. Wigner, 1959).

Perhaps we will find that just as no elementary descriptive systems such as nucleic acids and their transcribing and constructing enzymes are designed to directly formulate high-level symbolic concepts such as symmetry or conservation, so neither are our theories of simple physical systems designed to directly explain a description of even the most elementary symbolic processes. The genetic code as it now exists might suggest that there is a threshold of physical complexity below which no meaning can be attached to a description; that is, below this level we would only recognize physical interactions with no internal symbolic aspects. A similar suggestion was made by von Neumann (1966) in considering the logical design of a self-reproducing automaton which required an internal description. One can also see that some threshold of complexity of language is necessary to formulate what we would call even the simplest physical law. It is at this level of interplay between the most elementary descriptive activity and the physical systems that they constrain, that there may be some hope of discovering a deeper relation between the parts and wholes of hierarchies.

References

Bellman, R., & Kalaba, R. (1964). *Selected papers on mathematical trends in control theory*. New York: Dover Publications.

Glansdorff, P., & Prigogine, I. (1971). *Thermodynamic theory of structure. Stability and fluctuations*. London: Wiley-Interscience.

Pattee, H. H. (1971). The recognition of description and function in chemical reaction networks. In R. Buvet & C. Ponnamperuma (Eds.), *Molecular evolution* (p. 42). Amsterdam: North Holland Publishing Company.

Prigogine, I., & Nicolis, G. (1971). Biological order, structure and instabilities. *Quarterly Review of Biophysics, 4*, 107.

Rosen, R. (1970). *Dynamical systems theory in biology*. New York: Wiley-Interscience.

Thom, R. (1970). Topological models in biology. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Drafts, Vol. 3, p. 89). Edinburgh: University of Edinburgh Press.

von Neumann, J. (1966). In A. W. Burks (Ed.), *Theory of self reproducing automata*. Urbana: University of Illinois Press.

Whyte, L. L. (1969). Structural hierarchies: A challenging class of biological and physical problems. In L. L. Whyte, A. G. Wilson, & D. Wilson (Eds.), *Hierarchical structures* (p. 3). New York: American Elsevier.

Wigner, E. P. (1959). The unreasonable effectiveness of mathematics in the natural sciences. *Communications on Pure and Applied Math, 13*, 1.

Wilson, A. G. (1969). Hierarchical structures in the cosmos. In L. L. Whyte, A. G. Wilson, & D. Wilson (Eds.), *Hierarchical structures* (p. 113). New York: American Elsevier.

7 Discrete and Continuous Processes in Computers and Brains

7.1 Nerve Cells and Switches

Theories of computation and theories of the brain have close historical interrelations, the best-known examples being Turing's introspective use of the brain's operation as a model for his idealized computing machine (Turing 1936), McCulloch's and Pitts' use of ideal switching elements to model the brain (McCulloch and Pitts 1943), and von Neumann's comparison of the logic and physics of both brains and computers (von Neumann 1958).

The basis for this historical relationship, as well as for the vast literature on computers and the brain produced since then, is the assumption, for better or for worse, that the nerve cell functions something like a discrete switch. On the experimental side, this assumption is supported by anatomical and electrophysiological measurements showing that nerve cells form complex networks and communicate only by "all-or-nothing" discrete input and output pulses that can be related by a relatively simple, but modifiable, switching or transfer function. On the theoretical side, this assumption is supported by the universality of simple switching elements for representing any finite symbol manipulating process which can be translated into an effective algorithm or program.

Adding significantly to the credibility of the switch-neuron analogy are the specific observations of neural coding networks such as feature detector zones in the optical system, (e.g., Hubel and Wiesel 1968), whose general type of functional behavior can be simulated by artificial sensing arrays coupled to discrete computers, (e.g., Guzman 1968; Duda and Hart 1970). All this recent progress in neurobiology has led many to conclude that "the immense task of understanding the neural basis of perception is only immense; it is no longer incomprehensible" (Handler 1970).

Reprinted from *Physics and Mathematics of the Nervous System*, M. Conrad, W. Güttinger, Eds. Lecture Notes in Biomathematics 4. Berlin, Heidelberg, New York: Springer-Verlag, 1974, pp. 128–148.

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7, DOI 10.1007/978-94-007-5161-3_8,

Similarly we find many computer scientists who believe that, given enough switches, creating a truly intelligent, "thinking" computer is only a matter of programming these switches to behave in an intelligent way.

7.2 Are Discrete Switches Sufficient?

The question I want to raise is whether this concept of the discrete switch will prove sufficient as a basis for understanding intelligence in brains and creating intelligence in computers. There is no question that the analogy of the neuron and the discrete switch is fruitful and very likely sufficient for all permanently established functions or models in the brain, where no creative intelligence would serve any purpose; (e.g., in sensory and neuromuscular systems which interact directly with the real physical world). One has no need of learning to move or see in another physical world. I am only talking about intelligence as generation of new internal models. In other words, I am talking about how internal models of the world are first created and interpreted, not how they are finally represented or embodied. Once we have a well-defined model, it certainly can exist as a discrete structure (i.e. as a symbol structure), but I agree with Emil Post (1965) that symbols are *created* in continuous dynamical time, and are only *preserved* in discrete, arbitrary structures.

Therefore, I have no doubt that the discrete automaton can be used to represent or embody all well-defined sensory, neuromuscular or computational functions. In other words, when we can say exactly what we mean, then we are in the discrete symbolic mode, characteristic of the inputs and outputs of switching nets. This mode is programmable by discrete symbol systems, such as punched cards or tapes, and can approximate the behavior of many continuous dynamical systems. I expect, then, that this discrete symbolic mode is essential for clear descriptions and constructions and will continue to be one essential form of description for nervous systems. It is also clear that automata theory and real computers will continue to depend on formal and concrete switches respectively. My question is whether there are other important modes of behavior that cannot be adequately simulated by discrete events in automata. In other words, does all interpretation and intelligence involve only discrete events? Can we expect to account for, or simulate, the more intelligent problem-finding, pattern-interpreting or linguistic performance of the brain simply by more immense networks of discrete switches and discrete memories?

Now one might ask immediately, "What other type of description is there?" All concrete statements of our mathematical languages including continuous functions and variables are specified by a finite, discrete set of symbols and rules. Consequently, even a continuous dynamical system, such as the motion of several mass points in a potential field is "solved" in practice by approximating the values of the continuous variables over a discrete mesh, and representing the mesh behavior by an automaton. Furthermore, even our natural languages are made up of discrete, finite elements so that one could argue that all descriptions of continuous processes must be representable in some form by a finite, discrete sequence of finite elements.

One could also look at the phenomenological level of either the nerve cell or the switch and argue that the microscopic continuous dynamics is totally suppressed by the discrete functions of these units. Even cellular functions at the biochemical level appear to be adequately described by discrete sequences of specific reactions catalyzed by individual molecular subunits.

Thus both the automata theorist and molecular biologist provide very convincing evidence that discrete switching behavior has universality as an elementary concept for building abstract descriptions as well as a basic unit for building and controlling complex organisms.

To most physicists and mathematicians, on the other hand, the switch has never appeared as a fundamental unit, although the question of whether continuity or discreteness is more fundamental has always created problems and paradoxes. Von Neumann (1958) was one mathematician who took switches and neurons seriously. He said in the introduction to the *Computer and the* Brain:

> I suspect that a deeper mathematical study of the nervous system—'mathematical' in the sense outlined above [as a real computer]—will affect our understanding of the aspects of mathematics itself that are involved. In fact it may alter the way in which we look on mathematics and logic proper.

Furthermore, he did not regard computer theory as abstract automata theory, but as a theory of real, physical devices. He says in his *Theory of Self-Reproducing Automata* (von Neumann 1966):

> By axiomatizing automata in this manner one has thrown half of the problem out the window and it may be the more important half. One has resigned oneself not to explain how these parts are made up of real things, specifically, how these parts are made up of actual elementary particles, or even of higher chemical molecules. One does not ask the most intriguing, exciting, and important question of why molecules or aggregates which in nature really occur in these parts are the sort of things they are …

Von Neumann was well-aware of the basic principles that prevent physicists from regarding the machine, including the switch, as just a simple objective physical device (e.g., Pattee 1972). Many mathematicians and biologists still do not understand the reasons for this point of view. Even Turing (1936) in his article, Can a Machine Think?, and Gödel (1964) in equating "finite procedure" with "mechanical procedure" imply that relating thought or formal logic to a "machine" would constitute some fundamental reduction or explanation of higher mental processes.

7.3 A Physical View of a Switch

To a physicist, however, machines of all kinds, including switches, cannot be a fundamental explanation of anything. The physical parts of machines do, of course, obey laws of physics, but the concept "machine" is defined by its function, i.e., a machine performs useful work, or is a kind of prosthetic device for making man's work easier or simpler. The essential conditions are "useful," "easy," and "simple," and these are concepts that in no way can be explained by physical laws. In other words, a machine can only be defined by its constraints, not by equations of motion (cf. Polanyi 1968).

Insofar as a physical system can be recognized as a switch it is not an objective element, but is itself a model—a subjective interpretation made through some external system with which it interacts. There is no question that all switching behavior must possess an underlying physical dynamics, but the switching function itself cannot be an inherent property of this dynamics (Rosen 1969). In fact, it is the selective disregard of the detailed dynamics (selective dissipation), which must originate through an outside "selector," that encodes the dynamics into the switching function. It is this outside selective encoding that creates the switching behavior and suppresses the dynamical details. This outside agent has, in effect, created its own internal model of the dynamics, and it is this model which we recognize as the switch. The discrete switch and its continuous dynamics therefore comprise *complementary* descriptions of a two-level relation which cannot be understood by either description alone. This relation is essentially an "epistemological" relation of object-to-subject, matter-to-symbol, event-to-measurement or system-to-model, however you wish to express it.

7.4 A Philosophical View of a Switch

One could argue, on the other hand, that a switching brain is more fundamental than continuous dynamics because the concepts of continuous motion, and continuous space and time are really the abstract models created by the patterns of discrete switching activity of the brain. According to this point of view, the discrete switch is a fundamental objective element out of which is built the subjective interpretations of experience which includes the abstraction of continuity and hence physical equations of motion.

The logic of this somewhat solipsistic view appears sound, but I find this latter interpretation contrary to the more or less commonly accepted empirical facts of evolution. In particular, it presupposes the existence of brains as switching networks, whereas I find it more consistent with the historical view of the universe to presuppose the existence of space, time and matter. In other words, I find evidence which associates switching functions with living systems in an essential way, and I also find evidence that life did not exist on earth 5 billion years ago. Therefore I regard the origin and evolution of discrete switches as an evolutionary step that may be explained beginning from our dynamical view of non-living matter. I also find it natural to think of switching descriptions as a simplification or abstraction of the underlying dynamics, whereas I cannot conceive of the relation the other way around.

Before going into more detail, let me restate my general argument why discrete automata may be insufficient to generate intelligent behavior in brains or computers. By intelligent behavior, I mean the ability to create internal models of the world, on which predictive decisions are based. I see the discrete switch function as the primary *result* of this creative process, not the cause of it. The essential act of intelligence is the abstracted, simplified encoding of continuous, real-time dynamical trajectories into a discrete, switching *function*. Discrete programs instructing these discrete switches can, therefore, function as models of continuous dynamical systems. However, their function as models is not an inherent property of these automata, but requires an interpretation to relate the automata to the dynamical system. For this reason, I would not expect a formal switching net, nor a real switching net with "hard" switches (i.e., isolated from their dynamical matrix) to be capable of recognizing or interpreting dynamical systems. This relating of discrete symbols to continuous dynamics is a crucial form of intelligence.

What I feel must be distinguished is the end *result* of intelligent activity, which is usually a formal symbolic structure capable of being preserved by a discrete switching network or "mechanical device," and the intelligent activity itself—the *creation* and *interpretation* of these symbolic structures.

7.5 An Operational View of a Switch

To make these general ideas clearer I shall take the simplest example of a switch that I can imagine. I shall define it operationally then represent it as a formal, discrete function, and then describe its simplest underlying dynamics as far as that is possible. Next, I consider the necessary conditions for its physical embodiment and the stages of its evolution. Finally, I suggest biological examples at several levels of organization where I believe complementary discrete and continuous descriptions are necessary to understand the creative aspects of both evolution and intelligence.

What are the operational characteristics of a switch? First, it must have at least one time-independent stable position or state. Of course, this "time independence" is only with respect to the switching variable, which in macroscopic switches may be better described as a stable equilibrium. Second, there must be some selective degree of freedom, which when perturbed by an external force, called the trigger, results in a relatively sudden change of state called the action. The *threshold*, which depends on the sensitivity and selectivity of the switch to the trigger perturbation, is a crucial aspect of the interpretation or *function* of the switch. Typically, a good switch will not respond to perturbations below a certain threshold which is well above noise, while on the other hand, the strength of the trigger is less than the strength of the action produced. Also, perturbations of the nonselective degrees of freedom must not trigger the action. Finally, the switch should be useful more than once, which means it must either reset itself or be resettable to the original time-independent state by an external force. It is implied, of course, that there is a definite "rule" of operation, that is, that all previous or subsequent triggers would produce the same action. A switch function is summarized operationally, then, by the following:

1. At least one time-independent stable configuration.
2. A selective, sensitive, but reliable trigger which causes a sudden change of state or action.
3. A return to the initial state—reset, repeatability.

7.6 A Discrete Logical View of a Switch

The formal behavior of a switch might be abstracted from this description by calling the resting state, S_0 the action, S_1 the absence of a trigger, 0, and the presence of a trigger, 1. Then one particular switch is defined abstractly by the following Table 7.1:

This switch resets itself since if it is in the action state S_1 it returns to S_0 with or without a trigger. Therefore, during the time it takes the switch to reset, it is not sensitive to the trigger.

Next we might ask what the physical embodiment of such a discrete, formal switching function could be. Clearly this is not itself a formal question. There are innumerable physical systems which can be inter*preted* as functioning in this way, including mechanical, electrical, hydraulic, or molecular devices of all sizes and shapes. So this is not a very useful question as it stands. But what is it we would like to know beyond the formal switching table? Is the formal discrete abstraction the only universal description of this switch? Are all other more detailed descriptions arbitrary and hence extraneous to the function? This is the important question. If we are using the language of formal logic or computers we find that all details lose their significance below the functional level. How a truth table is actually executed, or how a switch is actually constructed simply has no bearing on the outcome. But the situation is more drastic than this. If in fact a formal switching function were not completely independent of its detailed embodiment, that is, if we had to know some additional details before we could define switching behavior, then we would either call it a bad switch, or we would say that we have misrepresented it (i.e., it really has a different switching function than what we thought).

What we have, in fact, done in writing this formal description of a switch is to define what we are going to mean by *input* and *output*, *trigger* and state. We have created this switch by this definition alone, without reference to any physical device. If any real dynamical system behaved like this switch it is because of our *interpretation* relating selected physical observables to the symbols in the formal definition. But the formal switch by itself has no underlying dynamics, hence the formal switching function is independent of dynamics *by definition*.

On the other hand, it is a meaningful question to ask what kind of continuous dynamical system could be reasonably interpreted as such a switch. What are the general conditions which allow a continuous real time dynamical system to be interpreted as a formal switch with a particular switching function, such as given by Table 7.1?

Table 7.1 A formal symbolic description of a switch

		State		
		S_0	S_1	
Trigger	0	S_0	S_0	Next state
	1	S_1	S_0	

7.7 A Continuous Dynamical View of a Switch

It is significant that the oldest dynamical descriptions of switching behavior (Blair 1932; Rashevsky 1933; Hill 1936) are in fact models of nerve cell excitation. These are phenomenological models with no direct attempt to describe the underlying chemical variables as in the Hodgkin-Huxley equations. More recently E.C. Zeeman (1972a), at the suggestion of Francis Crick, applied the theory of the cusp catastrophe of Rene Thom (1970) to derive the simplest dynamical description of switching behavior, again, using the nerve cell as one embodiment of the switch function.

It is also significant that these continuous dynamical models of discrete behavior of nerve cells were motivated by the assumption that the discrete switching function is the basic element of all nervous activity and that once this function is successfully modeled by a dynamical system, then one can dispense with the dynamics in all higher level models of nervous function. In other words, the problem in these treatments is limited to the explanation of how continuous dynamical processes that occur in the soft chemical cell can result in the discrete functions of a hard switch. To the extent that the brain functions like a discrete switching network, these dynamical models of switches are a useful reductionist description, and this is what they were intended to be. Our aim, on the other hand, is to suggest why the discrete switching description, while simpler and often more practical than the continuous dynamical description, cannot stand alone as the basis for its own creation or its own interpretation.

Zeeman (1972a) has presented the simplest possible continuous dynamical system that can be interpreted as a switch, satisfying both the operational conditions described in Sect. 7.5 and the logical switching function given in Sect. 7.6. To appreciate in what sense this is the simplest dynamical system with these switching properties and its mathematical relation to Thom's canonical cusp catastrophe, one must read Zeeman's paper.

The system consists of a "fast foliation" equation, $\varepsilon dx/dt = -(x^3 + ax + b)$, and the two parameters a and b are interpreted as coordinates, x, a, b forming a 3-dimensional space. A "slow manifold" is formed by giving a and b dynamics, $da/dt = -2a - 2x$ and $db/dt = -a - 1$. This slow manifold forms the folded surface with a cusp (Fig. 7.1) which is a single-sheeted attractor except for the three-sheeted region inside the cusp with the middle sheet a repellor. The interpretation is as follows:

The equilibrium or time-independent stable state is given by $dx/dt = da/dt = db/dt = 0$. A trigger is an external force moving b over the edge of the top attractor sheet. This edge is the threshold beyond which the fast foliation equation produces a sudden action, since the middle sheet inside the cusp is a repellor. When the trajectory reaches the bottom attractor, it returns slowly to the initial state by winding up around the cusp, resetting the switch

The formal interpretation requires a definition of the states S_0 and S_1. These states are the "resting" state and the "action" state, so we may simply divide the slow manifold into two regions such that one, S_1, contains the upper edge of the cusp where the action takes place, its boundary being such that all trajectories leaving

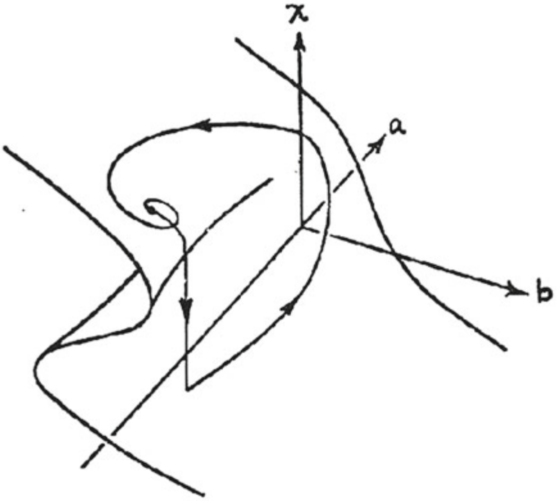

Fig. 7.1 A continuous dynamical model of a switch

S_1 enter S_0 and the other region, S_0, containing the equilibrium point, its boundary being such that no unperturbed trajectory leaves that region. A trigger is any displacement leading from the S_0 region across the upper edge of the cusp. This leaves a large arbitrariness in the shape of the two regions corresponding to S_0 and S_1. Perhaps more important is the arbitrariness in the details of the dynamical equations that can be interpreted as such a switch. The basic requirements are the region of equilibrium, the cusp topology, and the reasonably simple, fast foliation and slow manifold trajectories. This arbitrariness is important for any practical, reliable switch which must be relatively insensitive to design or construction variations, as well as external perturbations. This approximates the ideal switch which is sensitive only to the trigger.

7.8 A Dissipative View of a Switch

The mention of an "ideal switch" raises fundamental physical and logical problems. We have explained how a continuous dynamical system can be interpreted as a discrete switch with a logical behavior as described in Sect. 7.6. Now we must go one step further and ask, "Can any real physical system be described by the dynamical equations of Zeeman that we have been interpreting?"

Zeeman himself describes a very simple catastrophe machine as an appendix to his mathematical description (Zeeman 1972b). It consists of a disk with a fixed pivot, with two elastic bands attached at one point of the disk's circumference.

The other end of one elastic band is fixed two disk-diameters away from the pivot. As the other band is pulled over a two dimensional surface, parallel to the disk, the switching behavior of the disk will generate the cusp.

If one constructs this machine with great care, using fine bearings and elastic springs, so that there is very little friction, it will not return to equilibrium very quickly after being triggered but will undergo damped oscillations. Under these conditions, we must restrict the interpretation of S_0 to a region very close to the equilibrium point, otherwise the trigger threshold will be phase or time dependent. In other words, the switch will be sensitive to a great variety of input triggers that are not independent of the detailed dynamics. Such a device that does not effectively suppress the continuous dynamics would not be interpreted as a reliable switch.

This illustrates the fundamental physical condition imposed on all switches, or more generally, on all logical operations, that for every discrete switching event (or equivalently, for any record, measurement, classification or decision process) there must be a corresponding dissipation of energy. The absolute minimum cost is kT of dissipation for each binary switching event or bit: but all real switches, including neurons, dissipate orders of magnitude more energy. This is to assure greater reliability. It also turns out that dissipation increases with the speed of the switch (e.g. Landauer 1961; Brillouin 1962; Keyes 1970).

Therefore, the concept of the ideal "mechanical device" or "formal procedure" is no more realizable than the perpetual motion machine, and it is a serious physical and logical error to base any evolutionary theory of biological function only on such ideal elements, since natural selection does not operate on ideal switching functions, but on the real speed and reliability of the decision making dynamics. The same type of selection would also appear necessary for intelligent learning.

The only condition where the concept of the ideal switch is sufficient is in an ideal universe where space, time, matter and energy play no direct role. The most remarkable, and indeed, the most frightening aspect of modern computer technology is how closely it has approximated such an unnatural universe. However, one should constantly bear in mind that the only link between a computer and the real world (other than through the power mains and the air conditioner) is with the brain of a human programmer; and this link is accomplished only with alternative modes of description that presuppose a high level of creative intelligence on the side of the programmer. Yet we know that the practical and universal behavior of computers depends on this isolation of its hardware from everything but the programmer. The discrete mode of symbolic description which is characteristic of formal automata theory therefore appears to be necessary for precise description.

In my opinion, the trouble with discrete automata models of intelligence is that theorists have mistaken this necessity of isolation for its sufficiency. The interpretive or dynamical mode, since it is in the brain of the programmer, is left out.

7.9 An Evolutionary View of a Switch

What has happened with the evolution of computers is that the discrete sequential and switching mode has been artificially selected as the only significant behavior. It is then up to the programmer to translate every other mode of behavior into the particular discrete mode of the computer. The discrete symbolic mode in artificial computers has, in effect, been lumped into one box called the hardware where no continuous dynamical interactions or interpretation is allowed. Only as programs are written or as the output is read can the activity be interpreted or interact with the physical world.

Living systems also keep their discrete switching modes separate, but do not lump them all together. Instead, we find the discrete modes interspersed with dynamical modes at many levels. In fact, an essential requirement of evolution is the clear separation of the discrete genetic instructions from the phenotypic constructions. But this discrete genetic mode is also interpreted dynamically even at the subcellular level. For example, the cell's enzymes are constructed by a discrete code that translates the sequence of DNA bases to a sequence of amino acids; but the linear discrete string of amino acids does not function as a selective catalyst until it folds up into a three dimensional active form. This folding is not programmed, and can only be described by continuous dynamical interactions of the entire string. Going one stage further, the recognition of the substrate by the enzyme also requires a dynamical description with many degrees of freedom, but once the substrate is bound there follows the sudden action of the catalytic event, which has all the characteristics of the discrete switch we have been discussing.

This complementary alternation and interaction of discrete symbolic modes with the continuous dynamical modes goes on at all levels of biological organization from the enzyme structure-function levels up through the sensory and neuromuscular structure-function levels. One is inclined then to ask if this complementary discrete and continuous interaction is not also essential for the higher forms of intelligent behavior.

In particular, since the creative aspects of evolution are our best analogy to creative intelligence, we should ask if the process of evolution illustrates some general principles relating discrete and continuous modes. Indeed, a fundamental fact of Darwinian evolution is that the mutation or search process begins in the discrete symbolic mode, but that the interpretation of these symbols (i.e., selection) is a dynamical interaction in real space and time. For example, consider the enzyme again. Its evolution requires a discrete change in the linear sequence of its genetic *description*, not in the enzyme itself. There is no real-time or continuous process, such as the folding of its DNA description or *rate of reading* the sequence that has any bearing on this discrete mode. The genetic hardware is well isolated from real-time dynamics. On the other hand, whether or not such a change in description .is selected depends first on how the new enzyme folds up. The folding process, as we said, is not a part of the enzyme's description but is a real-time continuous dynamical interaction involving many degrees of freedom. Of course, this folding is only one step in the selection process, but we may also call it one step in the interpretation of its description, the translation by the genetic code being the first interpretive step. At the next functional level this same enzyme may be interpreted further as a discrete switch in a network of catalyzed reactions and hence may also be selected on the speed, specificity and reliability of its switching function in this network (e.g., Kauffman 1969).

Thus the evolutionary view of switching function is complementary to the physical view. We see that *from the physical view the switch function is a discrete interpretation of an underlying continuous dynamical system. From the evolutionary view, on the other hand, we see that the selection process is a dynamical interpretation of the underlying discrete switch-like genetic descriptions*. Zeeman's description of the nerve pulse is one example of how continuous dynamics creates discrete activity, and the synthesis of the enzyme is one example of how discrete activity constrains the continuous dynamics of folding. This folding in turn creates the dynamical interactions that result in the switch-like selective catalysis characteristic of enzyme function at the next level.

7.10 Dynamics, Language and Intelligence

These few simple examples illustrate very general properties of discrete systems and their relation to continuous systems. As we pointed out in Sect. 7.2, all types of descriptions from the gene to our highly evolved languages are represented by sequential elements and sequential operations. Formal automata theory and real computers represent extreme efforts toward isolation from real dynamical behavior. Since formal automata processes are performed in the brain as well as in machines there must be switching behavior in the brain which also appears isolated from real-time dynamics.

On the other hand, we have argued that switching behavior, while suppressing continuous dynamical interactions cannot be entirely isolated from dissipative fluctuations and even more important, cannot originate or have any evolutionary significance without interfacing with lower and higher levels of dynamical interactions. Specifically, we associate the *origin* of switching behavior with singularities in a continuous dynamical system and the significance or interpretation of switching behavior as a dynamical "folding" process that is essentially time-dependent and continuous. This interpretive dynamical mode may itself exhibit singularities that generate a new level of discrete description, and so this process may be repeated, producing a hierarchy of discrete symbolic levels interfacing on the lower generative side as singularities of a continuous matrix and on the upper side as constraints within a larger dynamical system.

We have used only the most primitive levels of biological organization, the gene and enzyme levels as examples. How would this discrete-continuous interplay apply to the highest levels, that is, to intelligence? How do discrete and continuous interactions apply to the higher levels of linguistic performance, for example? At this level we have very much less knowledge than we have of the molecular levels, and almost all of this knowledge is derived from the discrete mode of symbol manipulation and the transformation or coding to other discrete symbol structures. In other words, most of our knowledge of language is based on observations at the syntactical level. When we look on either side of this level, that is, at the origin and generation of language on the one side, or the interpretation of language on the other, our knowledge becomes vague or non-existent.

What little we do know of the creative level, however, does not support any theory based on the manipulation of well-defined discrete symbols. The creative act in all fields, according to the introspection of those who create, appears as a sudden flash, coherence, or harmony of indistinct psychical entities that, at first, cannot consciously be recognized or manipulated by the creator at the discrete verbal level.

Only after much effort can these creations be transformed into formal expressions. We recall the well-known letter of Einstein to Hadamard (1945) where he says, “The words or the language as they are written or spoken do not seem to play any role in my mechanism of thought,” and Einstein describes the pre-linguistic entities with which he plays as “of visual or muscular type.” Poincaré (1913) was also clear about the essential role of the non-logical level that generates the “sudden illumination” only after an extended period of “non-mechanical” activity. The recognition of “harmoniously” disposed elements Poincaré attributes to the “delicate sieve” of “esthetic sensibility.”

Although these examples can be multiplied over and over, and are characteristic of creation in music, literature and art as well as physics and mathematics (e.g., Ghiselin 1952), one cannot prove from introspection alone that the underlying unconscious level is not some kind of discrete network. All we can say is that taken at their literal value, these descriptions of the creative process suggest a picture of singularities arising in a dynamical sea of ideas rather than the output of a discrete automaton.

On the other side of the discrete symbol level of language—the reading and interpreting of statements—we know that translation of such discrete symbols most often results in the performance of some dynamical action. For example, when we see the muscles actually execute a task according to instructions, we know that some discrete pulse patterns have been converted through an intricate set of constraints to a continuous dynamical action. We have evidence that the message in this case lies in the statistical distribution of discrete pulses and not in the detailed pulse timing, and this leads to continuous dynamical models of nerve pulse interpretation depending on our choices of how pulses are averaged (e.g. Griffith 1963; Harth et al. 1970; Anninos et al. 1970; Cowan 1973).

However, as we have said, motor control is relatively simple and fixed. What we know of the interpretation of higher levels of language does not fit either a discrete automaton model or the statistical dynamical models with only simple averaging. Our present automata and statistical models are actually extreme cases of interpretation themselves. At the automata extreme we assume that each elementary symbol is completely interpreted by its input, output or state specification as we indicated in Sect. 7.6. One, therefore, cannot discover new observables within such automata (Rosen 1969).

On the other hand, to arrive at any statistical dynamics of a discrete system also requires a definite procedure for generating observables from the discrete elements. The other extreme case is when all elements are lumped together by a fixed and uniform averaging process that is entirely independent of the sequence of elements.

We know that the interpretation of natural language is not this simple and cannot be explained or simulated by either of these extreme models. The elements of our language, the words, cannot be represented as fixed elements of a discrete automaton or as the observables of a fixed, uniform averaging process. The way we interpret strings of words appears to be much closer to the way the cell interprets strings of DNA, where the meaning of the discrete elements is first created by a highly context-sensitive folding of these elements into functional or meaningful units.

Furthermore, we know that interpretation does not stop at one level either for a single sentence or a single enzyme. For example, even if an enzyme is synthesized and folds up locally in the proper way to function, it is likely that this is only one enzyme in a string of enzymes where the end-product is the global function for which this string has been selected. Furthermore, the control of the organism as a whole may require that this end-product act like a switch, inhibiting the reading of the genes producing this string.

In a similar sense, a sentence such as “Turn left at the second stop sign,” may be interpreted quite clearly at the local level, but may be only one of a string of sentences which has a more global function of taking you, say, to a gasoline station. Finally, to complete the analogy, the meaning of the entire string of sentences may be overridden by the sentence, “But I do not need gasoline.” We do not know, of course, that this is a good analogy, since we have almost no evidence about how natural language is interpreted in the brain. However, we know enough about the folding and control of enzymes to know that both discrete and continuous descriptions are necessary for modeling their behavior.

One point needs to be clarified here. The claim that it would be possible to describe the folding process by a discrete program in a computer does not mean that continuous dynamical processes are not essential for the origin or interpretation of symbols. The reason for this, as we have noted, is that the process of evolution occurs in real dynamical time (i.e., natural selection is not a symbolic process), and therefore, the speed and reliability of interpretation is crucial. It follows that a discrete symbolic representation of a dynamical process is selectively competitive only if it can *predict* results faster than the result itself occurs. In the case of the enzyme, the evidence is very strong that the folding is not a programmed or discrete process, and no one has even imagined a predictive simulation which would take less time.

7.11 Brains, Computers and Intelligence

We have proposed that the hierarchical levels of organization in living systems are based on an interaction of both continuous dynamical modes and discrete switching modes at each level. We associate the discrete switching modes with the *informational*, *programmable*, or *linguistic* functions, such as the linear sequences of nucleotides in a single DNA molecule or the sequences of pulses generated by a single neuron. We associate the singularities of continuous dynamical modes with the *creation* of these discrete sequences and the parallel dynamical interactions or "folding" of these sequences with their *interpretation*. This continuous model is not effectively programmable, but results from the non-integrable physical constraints that couple the rate dependent sequential events of the discrete mode with the continuous real-time dynamics of the system.

We have suggested that this same type of complementary interplay of continuous and discrete activity is also the key to intelligent behavior, since intelligence implies not only a symbolic representation or model of the world, but even more, the *creation* and *interpretation* of this model. What evidence do we have from our study of the brain that supports this suggestion? As we said at the beginning, there is not the slightest doubt that the discrete, all-or-none, pulse mode of activity exists and can account for many signaling and data-processing functions of the brain. The only doubtful assumption is that this mode alone is adequate for all forms of memory, learning and intelligent behavior.

The important part of the question, then, is what types of continuous dynamical behavior can we find in the brain that can be associated with the creation and interpretation of patterns of these discrete pulses? I shall mention only two possibilities, one at the molecular genetic level, and one at the whole brain level of organization.

At the molecular genetic level, there is growing evidence supporting Sperry's (1970) hypothesis that the nerve net circuits over which the pulses travel are constructed by utilizing the same type of detailed chemical specificity that we associate with enzyme-substrate recognition. In the case of the nerve cell growth, the process is undoubtedly even more complex, involving the interactions of highly specific membrane structures, and perhaps microfibrils or microtubules. In spite of our ignorance of the many intricate details of this growth and specificity, it is reasonable only from our present knowledge that several levels of discrete and continuous interaction processes are involved.

What does not seem reasonable is that the growth of the brain is only a discrete switching process, and it is equally unreasonable to expect that such a highly evolved level of continuous chemical dynamics, growth and morphogenesis should at a later stage be entirely suppressed in favor of the discrete switching mode. Indeed the chemical evidence now makes the idea that all signaling in the brain is by all-or-none nerve impulses untenable, and as Eccles (1973) says, "We can now postulate that the whole nervous system has communication not only by impulses . . . but also chemically by transport of specific proteins or other macromolecules." Since the only known mechanism of specific recognition involves the "folding" of strong-bonded sequences by highly parallel dynamical interactions of weak bonds, we may reasonably postulate that this type of continuous dynamics plays an essential and complementary role in the interpretation and generation of the discrete nerve pulses at many levels (Conrad 1974).

As a second possible example of the complementary interfacing of discrete and continuous modes, we may look at the highest level of brain organization, and consider the functions of the so-called dominant and minor cerebral hemispheres. Of course, at this level we are observing only the broadest general functions. We have no idea how many intervening levels of generative, symbolic or interpretive modes exist. We have no idea how many times the discrete pulse patterns at one level have been coded into a statistical dynamical continuum whose singularities have in turn generated a new discrete switching mode at a higher level. In any case, at the functional level the dominant hemisphere has been characterized by its verbal, analytic, arithmetical and sequential activities, while the minor hemisphere has been characterized by its musical, synthetic, geometrical and holistic activities (Sperry 1970; Eccles 1973). I would suggest simply that at the level we observe them the dominant hemisphere is operating primarily in the discrete switching mode while the minor hemisphere is operating primarily in a continuous dynamical mode. I would also infer, therefore, that the minor hemisphere is primarily generative and interpretive with respect to the discrete symbolic and linguistic activities of the dominant hemisphere.

One very significant fact should be borne in mind, that in spite of some degree of gross anatomical difference between the two hemispheres, there is no evidence of any basic difference in the nerve cells or the way in which they grow or are connected. This may appear puzzling if one assumes that discrete or continuous behavior should be observable as a structural property of the brain. But one point that I am trying to emphasize is that discrete and continuous modes are not intrinsic physical or logical properties of structures, but are themselves an interpretation of how these structures interact with each other. Again I return to the enzyme as the simplest paradigm example. Whether the enzyme is interpreted as a discrete or continuous device depends on which interactions are significant at any given time. The discrete, linear sequence of amino acids is only distinguishable from the continuous, three-dimensional folding because of the distinction between strong and weak bond interactions, since the strong bonds effectively define the discrete linear sequence while the weak bonds constrain the continuous dynamics of substrate recognition and catalysis.

In a more general way, Thom's catastrophe models, as exemplified here by Zeeman's dynamical view of a nerve pulse (Sect. 7.7), show how the discrete and continuous modes may be related. Here we interpret the discrete switching function as the "fast foliation" at the cusp singularity (the nerve pulse), which can exist only because of the continuous dynamics of the "slow manifold" in higher dimensions (the biochemical matrix).

Bear in mind that these are only the simplest conceivable examples illustrating the complementary relation of discrete and continuous modes. The enzyme molecule is, after all, only a functional unit within the context of an immensely more complex cell, and Zeeman's dynamical model of a switch was explicitly derived as the simplest possible representation of a cusp catastrophe with given switching characteristics. There is, therefore, little justification for assuming that any activity of the nervous system could be realistically simulated with model neurons of greater simplicity.

Finally, this brings me to the questions of why computers are not more like brains, and how we can design computers to be more intelligent. If what I have said about the generative, symbolic and interpretive roles of continuous and discrete modes has some degree of truth, then the answer to the first question is quite clear: Even the largest imaginable computer, if restricted only to the discrete switching mode of present day computers, can at best approximate only half a brain, and this will be the sequential, analytic half, not the generative, interpretive half.

Even in comparison with enzyme function, present computers appear impoverished; for they have no natural equivalent of strong and weak interactions, hence no sequence folding operations, no simple procedure of pattern or substrate recognition and no corresponding mechanisms for selectively catalyzing the *rates of growth* of their own sequences. While these inherently continuous dynamical processes can be programmed to some degree, it is fair to say that what small successes have been obtained are the result of very large programs generated by the brains of very intelligent programmers.

A common remark intended to account for the functional discrepancy of brains and computers is that brains perform "parallel" computations while computers are, so far, only "sequential" machines. This may be part of the problem, but as long as we mean by "parallel" only more simultaneous *discrete* operations, I do not think it is the basic problem. As I pointed out in Sect. 7.9 on the evolution of computers, the discrete switching mode has been artificially selected as significant while all continuous dynamical interactions have been artificially suppressed. Consequently, the generation of new observables or discrete elements as singularities in an underlying continuum, and the self-interpretation of linear sequences by folding processes are precluded.

The answer to my second question of how to design more intelligent computers is very simple to state, but very difficult to implement. As we have argued, the problem is not in the programs or the organization of the switching elements, but in the limited functions of the elements themselves. By abstracting the switch too far we have, as van Neumann said, "thrown half the problem out of the window and it may be the more important half." The trick will be to learn how to reintroduce a continuous dynamical mode into an artificial computer element at a simple enough level to be practical. To do this, I believe we must first have simpler, clearer theoretical models and interpretations of "intelligent behavior."

Acknowledgement This work was supported in part by grant from National Aeronautics and Space Administration, No. NGR 33-015-002.

References

Anninos, P. A., Beek, B., Csermely, T. J., Harth, E. M., & Pertile, G. (1970). Dynamics of neural structures. *Journal of Theoretical Biology, 25*, 121–148.

Blair, E. (1932). On the intensity-time relations for stimulation by electric currents. *Journal of General Physiology, 12*, 709.

Brillouin, L. (1962). *Science and information theory*. New York: Academic.

Conrad, M. (1974). Molecular information processing in the central nervous system. Part I, selection circuits in the brain. In M. Conrad & W. Guttinger (Eds.), *Physics and mathematics of the nervous system*. Lecture notes in biomathematics 4 (pp. 82–107). Berlin/Heidelberg/New York: Springer.

Cowan, J. D. (1973). Stochastic models of neuroelectric activity. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Vol. 4, pp. 169–188). Edinburgh: University of Edinburgh Press.

Duda, R.O., & Hart, P.E. (1970). Experiment in scene analysis. In *Proceedings of the First National Symposium on Industrial Robots*, Chicago, April, 1970.

Eccles, J. C. (1973). *The understanding of the brain*. New York: McGraw-Hill.

Ghiselin, B. (Ed.). (1952). *The creative process*. Berkeley: University of California Press (Reprinted 1955 by Mentor Books, New York).

Gödel, K. (1964). On formally undecidable propositions of the *Principia Mathematica* and related systems. In M. Davis (Ed.), *The Undecidable* (pp. 4–38). Hewlett: Rowen Press.

Griffith, J.S. (1963). A field theory of neural nets: I. Derivation of field equations, *Bulletin of Mathematical Biophysics, 25*, 111; II. Properties of field equations, *Bulletin of Mathematical Biophysics, 27*, 187.

Guzman, A. (1968). Decomposition of a visual scene into three dimensional bodies. In *Proceeding of the Fall Joint Computer Conference,* Washington DC: IEEE Computer Society Press, 291–304

Hadamard, J. (1945). *The psychology of invention in the mathematical field*. Princeton: Princeton University Press.

Handler, P. (Ed.). (1970). *Biology and the future of man*. New York: Oxford University Press. 360.

Harth, E. M., Csermely, T. J., Beek, B., & Lindsay, R. D. (1970). Brain function and neural dynamics. *Journal of Theoretical Biology, 26*, 93–120.

Hill, A. V. (1936). Excitation and accommodation in nerve. *Proceedings of the Royal Society B of London, 19*, 305.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195*, 215–243.

Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed nets. *Journal of Theoretical Biology, 22*, 437–467.

Keyes, R. W. (1970). Power dissipation in information processing. *Science, 168*, 796–801.

Landauer, R. (1961). Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development, 5*, 183–191.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5*, 115–133.

Pattee, H. H. (1972). Physical problems of decision-making constraints. *International Journal of Neuroscience, 3*, 99–106.

Poincaré, H. (1913). Mathematical creation, from *The Foundations of Science* 1913, 1946. New York: Science Press

Polanyi, M. (1968). Life's irreducible structure. *Science, 160*, 1308–1312.

Post, E. (1965). Selections from diary of E. Post. In M. Davis (Ed.), *The undecidable* (p. 420). Hewlett: Rowen Press.

Rashevsky, N. (1933). Outline of a physiomathematical theory of excitation and inhibition. *Protoplasma, 20*, 42–56.

Rosen, R. (1969). Hierarchical organization in automata theoretical models of the nervous system. In K. N. Leibovic (Ed.), *Information processing in the nervous system* (pp. 21–35). New York: Springer.

Sperry, R. W. (1970). Cerebral dominance in perception. In A. Y. Francis et al. (Eds.), *Early experience in information processing in perceptual and reading disorders*. Washington, DC: National Academy of Sciences.

Thom, R. (1970). Topological models in biology. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Vol. 3, pp. 89–116). Edinburgh: Edinburgh University Press.

Turing, A.M. (1936). On computable numbers with an application to the Entscheidungs problem. In *Proceedings of the London Mathematical Society*, Series 2, 42, 230–265.
von Neumann, J. (1958). *The computer and the brain*. New Haven: Yale University Press.
von Neumann, J. (1966). *Theory of Self-Reproducing Automata.* A.W. Burks (Ed.), Urbana: University of Illinois Press
Zeeman, E. C. (1972a). Differential equations for the heartbeat and nerve impulse. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Vol. 4, pp. 8–67). Edinburgh: Edinburgh University Press.
Zeeman, E. C. (1972b). Appendix: A catastrophe machine. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Vol. 4, pp. 276–282). Edinburgh: Edinburgh University Press.

8 The Complementarity Principle in Biological and Social Structures

Abstract Complementarity is an epistemological principle derived from the subject-object or observer-system dichotomy, where each side requires a separate mode of description that is formally incompatible with and irreducible to the other, and where one mode of description alone does not provide comprehensive explanatory power. The classical physics paradigm, on which biological, social and psychological sciences are modeled, completely suppresses the observer or subject side of this dichotomy in order to claim unity and consistency in theory and objectivity in experimental observations. Quantum mechanical measurements have shown this paradigm to be untenable. Explanation of events requires both an objective, causal representation and a subjective, prescriptive representation that are complementary. The concepts of description and function in biological systems, and goals and policies in social systems, are found to have the same epistemological basis as the concept of measurement in physics. The concepts of rate-dependent and rate-independent processes are proposed as a necessary distinction for applying the principle of complementarity to explanations of physical, biological and social systems.

8.1 Introduction

The principle of complementarity is associated in a modern scientific context with Niels Bohr and the interpretation of quantum mechanics. It is still a controversial concept, not formalized nor explicitly defined, even in physics (Jammer 1974), and subject to a wide variety of generalizations and interpretations relating to the non-physical problems where its meaning is far removed from Bohr's original ideas. I would therefore consider it a bad risk to invoke complementarity in explanations of social and biological systems unless (a) there are overwhelming reasons to believe

---

Reprinted from *J. Social Biol. Structures*, 1:191–200, 1978.

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7, DOI 10.1007/978-94-007-5161-3_9, © Springer Science+Business Media Dordrecht 2012

that the concept is fundamental for explanations in a general epistemological sense, and (b) there are explicit rules for applying the concept to observable aspects of social and behavioral systems. This paper is an attempt to outline my reasons for believing that complementarity is a fundamental requirement for explanatory models of social and biological systems, and to give some rules for distinguishing observable complementary aspects of these systems.

There is, of course, very good reason to first question whether a principle that was apparently formulated to interpret the wave-particle duality of electrons should have any bearing on the immensely more complex functional systems that we recognize as living. The answer to this depends on whether complementarity arises from ontological or epistemological considerations, that is, whether the principle has more to do with the nature of electrons or with the nature of observers. Bohr (1928) was quite clear about this. He stated that the idea of complementarity '(…) bears a deep-going analogy to the general difficulty in the formation of human ideas, inherent in the distinction between subject and object.' Von Weizsäcker, who was one of Bohr's students, expressed the epistemological content of complementarity as entirely observer-dependent, where one aspect of measurement must reflect the passive mode of input to the observer's senses, while simultaneously the observer must actively choose which possible measurements are to be executed. Thus, von Weizsäcker (von Weizsäcker 1949) says: '…zwei Grundfunktionen des Bewustseins gehen in jeden Satz der Naturbeschreibung ein: Wissen und Wollen.' ['Two fundamental functions of consciousness underlie every statement in physics: cognition and volition.'] As Polanyi (Polanyi 1962) has pointed out so clearly, this duality is implicit in all observations, with the primary difference between physics and the biological and social sciences being an increased awareness of the volitional role of living systems when they are themselves the object of observation.

Bohr also attempted to apply the complementarity principle to biology (Bohr 1937) and psychology and human cultures (Bohr 1938), although in none of these papers is there any explicit rule or procedure suggested for associating an observable with a particular complementary mode of description. Bohr's aim was not to show that atoms are in any way analogous to living organization, but rather that in the process of observation there is 'an epistemological argument common to both fields'.

It is the epistemological argument that I wish to develop in this paper. What I shall attempt to show is that explanatory knowledge of biological and social systems—from cells to human societies—requires the simultaneous articulation of two, formally incompatible, modes of description. The source of this requirement lies in the subject-object duality, or the distinction between the image and the event, the knower and the known, the genotype and phenotype, the program and the hardware, or the policy and the implementation, however one may choose to express this basic distinction. The essence of the concept of complementarity is not in the recognition of this subject-object distinction, which is common to almost all epistemologies, but in the apparently paradoxical articulation of the two modes of knowing.

8.2 The Nature of Complementarity

It would not be useful here to review the complex and often controversial interpretations of complementarity. An excellent discussion of the history of the concept can be found in Jammer (1974). What I shall try to do is develop my own expressions of complementarity, keeping them as consistent as possible with Bohr's images, but emphasizing their applicability to explanations of social and biological systems. In many ways, it is more important to begin defining the concept in particular cases. The most severe obstacle to the assimilation of this spirit into our scientific explanations is the logically unresolvable incompatibility of the two complementary modes of description.

Western philosophy and science have placed the highest priorities on unity and consistency as criteria for explanations or theories, so that even the suggestion of a paradox or multiple representations of an experience induces a tension which we feel must be resolved at all costs. Whether or not the roots of this tension lie in theological or moral dogma, logical or psychological fear of contradiction, or biological or social competition, we do not accept with emotional neutrality two explanations of one and the same situation. One theory we feel must be more correct or of higher quality than the other. Furthermore, if one can find an interpretation of the two explanations with some overlap that appears incompatible, then we assume that one theory is right and the other wrong, and all our efforts are directed at determining which is which. Even today, the emergence of one dominant theory from the great historical controversies, such as the Copernican and Darwinian paradigms, is usually regarded as the inevitable victory of the right theory. It is a very rare scientist who will argue, like J.B.S. Haldane (1963), that Darwinian theory is about 80% correct and Lamarck about 20% correct.

There is no doubt that the theories of relativity and quantum mechanics increased the tolerance or acceptance of alternative points of view in scientists who had until then followed a hard-core paradigm of classical Laplacian determinism and the objectivity of experiments with the subsequent need for a decision on the correctness or incorrectness of any theory. More recently, the complexities of the social and psychological sciences have generated so many theories that there are increasingly frequent attempts at integration of alternatives, dialectical synthesis, transcendence of dichotomies and simply 'taking a broader view' as in the parable of the blind men and the elephant.

Unfortunately, none of these eclectic or tolerant views reflect the spirit of complementarity which emphasizes the necessity of formal incompatibilities in the dual modes of description, in contrast to the unity and consistency of the classical paradigm of a unified formalism. At the same time, however, this duality of descriptive modes and their incompatibility should not be thought of as a contradiction in any sense. In fact, there is none since the two modes of description are formally disjoint, and contradiction can only occur within a single formal system. Furthermore, this necessity of dual modes of description should not be interpreted as corresponding to a duality of the system under observation. In particular, it was characteristic of

Bohr's discussion of quantum systems to emphasize the wholeness or unity of the entire experimental situation (e.g. see Bohr 1949; Bohm 1971).

To further contrast the classical with the complementarity paradigm, we could say that classical explanation was achieved through the representation of conceptually disjoint systems by one, unified and coherent mode of description; whereas complementarity achieves explanation of conceptually unified coherent systems by disjoint modes of description. Thus we conceive of elementary systems such as stars, planets and ponderable objects that we recognize as separate entities with no functional relation between them, being unified by the law-of-motion mode of description. On the other hand, we also recognize complex systems that are conceptually unified by their functional coherence, such as machines, organisms and social structures, that are not adequately explained by any unified laws or analytical procedures. This is because we initially identify or define these complex systems of interest by their function rather than by their structure.

One point needs to be clarified. It will appear strange to some that complementarity should arise at the atomic level, apparently disappear at the classical macroscopic level and then reappear at the biological and social levels of organization. If functional activity is associated with the necessity of complementary modes of description, how do we account for any function in atomic systems?

Or, alternatively, why are macrophysical systems the only ones that are exempt from functional behavior? I believe the reason for this lies in the fact that macrophysical systems are the only level where the role of the observer, or the subject in the subject-object duality, can be suppressed and we still obtain reasonable predictions. As von Weizsäcker stated, every statement in physics involves cognition and volition on the part of the observer. At the macroscopic level we have gone to great effort to suppress the volitional aspect of measurement, although it is still there as Polanyi makes clear; and because celestial objects are exceptionally well isolated from the control of observers this suppression succeeded only too well, and is now applied everywhere—even to the extreme of trying to explain cognition and volition by unified, mechanistic models. In quantum theory, however, the evidence is clear that the measuring device and the atomic particle being measured must be interpreted as a unified, functional system described both by causal inexorable laws in one mode and as volitional measurement constraints in the other mode. I shall now proceed to outline the evidence for my claim that explanatory theories of biological and social systems have the same epistemological necessity for complementary modes of description—one mode representing the laws of the system as object and the other representing the constraints of the system as subject.

8.3 Generalized Complementarity

In order to understand this epistemological basis of complementarity I believe we must make a clear distinction between laws of nature and rules of constraints. It is a universal property of language (and hence, all descriptions) that the structure of symbol vehicles or signs (i.e. the letters of the alphabet, nucleotides, words, codons, etc.)

are related to their referent or their effect by arbitrary rules. These rules are not derived from, or reducible to, the laws of nature. They are perhaps best described as frozen historical accidents—accidents because their antecedent events are unobservable, historical because the crucial events occurred only once, and frozen because the result persists as a coherent, stable, hereditary constraint. The basic distinction between laws and rules can be made by these criteria: laws are (a) *inexorable*, (b) *incorporeal* and (c) *universal*; rules are (a) *arbitrary*, (b) *structure-dependent* and (c) *local.* In other words, we can never alter or evade laws of nature; we can always evade or change rules. Laws of nature do not need embodiments or structures to execute them; rules must have a real physical structure or constraint if they are to be executed. Finally, laws hold at all times and all places; rules only exist when and where there are physical structures to execute them.

A description is a rate-independent representation of certain constraints of the system where the structure of the description is related to the constraints by a coherent set of rules. This coherent set of rules is called the syntax of the description language. The rate-dependent, dynamical effects of these constraints may be called the semantics or the meaning of the description. This concept of language implies a threshold of complexity that has never been clearly defined, but we may take the universal genetic language as one of the simplest paradigm examples. In all known living organisms there is a genetic description in the form of the DNA molecule, which is the symbol vehicle, just as the ink on this page is the symbol vehicle for my language. The description of the physical state of the DNA, even at the detail of quantum mechanics, would give no more clue to the meaning of this string than the chemistry of this ink would give a clue to the meaning of my words. The meaning of the genetic DNA is the result of the coding and construction rules of the cell that are executed by hundreds of highly coordinated enzymes and nucleic acids that are themselves described by the genetic DNA. The details of this self-description, self-construction system are quite well known, but what I want to discuss are the system properties that require complementary modes of description.

At the biological level, why is it that the most detailed description of DNA in the physical mode will give no interpretation of what the DNA means? Why is it that a complete description of the informational sequence of DNA along with all the details of the coding give no clue to the laws under which all these structures and mechanisms operate?

At the social level, why is it that detailed description of behavior alone does not allow us to derive the policy underlying this behavior? And why is it that a complete statement of policy does not allow us to derive the forces and dynamics under which these policies will operate?

8.4 Rate-Dependent Laws, Rate-Independent Rules

The functional reason for the irreducibility of either laws to rules or rules to laws is that their descriptive modes are incompatible with respect to rates of events. All of the physical laws of motion are expressible as functions of rates, that is, as derivatives of some variables with respect to time. Time in physics is an irreducible concept, but by no means arbitrary. It is one member of the coherent set of concepts, including space, momentum and energy, that forms the language of physical laws and it should not be confused with timing intervals, the ordering of events or psychological time. The concept of rate, however, may be extended to social dynamics or psychological events.

Rules, on the other hand, depend on order but have no rate dependency and cannot be expressed as functions of a derivative with respect to physical or other time scales. All linguistic operations and all computations, in so far as they are defined by rules, cannot be functions of the rate of writing, reading or computing. That is, what you mean by a statement of policy or the value of a calculation cannot depend on how fast you speak or calculate.

Now one might argue that since all rules require physical structures for their execution, and since all structures must obey physical laws, it follows that the rules must be predictable from the laws. While the premises are correct, the conclusion is not, because rules are not interpreted in the same language as laws. For example, there is no doubt that logical rules can be executed by switching networks and that switching networks obey physical laws. But it does not follow that the rules of logic are predictable from the laws. Of course we also know that formal rule systems cannot derive the laws of nature. This fundamental complementarity was succinctly expressed by Einstein: ‘In so far as the propositions of mathematics [rules] are certain they do not refer to reality [laws], and in so far as they refer to reality they are not certain.’

8.5 Complementarity Applied to Biological Systems

A classical complementarity relation in biology is the structure-function duality. One may even paraphrase Einstein and say that in so far as structural descriptions are accurate [objectively] they do not refer to function, and in so far as functional descriptions are accurate [subjectively] they do not refer to structure. Explanation of biological systems must describe both structure and function. As we said earlier, the most accurate calculation of the DNA structure would have no necessary relation to its function as a template or description of an enzyme, nor would a precise description of these functional activities allow us to derive its quantum mechanical structure. Functional description is intrinsically a subjective concept in the sense that it is a property of the organism as a whole and even though each element of the organism obeys all the causal laws (i.e. the rate-dependent dynamics) of the objective description, the function or goal of these combined elements remains ineffable in this causal, objective mode of description.

The enormous success of modern molecular biology, in attaining detailed structural descriptions of living systems, temporarily eclipsed the problems of explaining the functional processes of self-description, self-construction and self-control that are characteristic of all living systems at all levels of organization. Very few modern biologists are willing to discuss living systems in functional or teleological languages, since this is contrary to the classical physics paradigm by which they currently measure explanatory power. On the other hand, those few biologists who also see the problem of living organisms as subjective teleological systems do not understand the epistemological basis of complementarity. For example, Monod (1971) viewed living systems as essentially teleological and saw no way in which the causal, objective mode of description could be modified to resolve what he called a 'flagrant epistemological contradiction'. Indeed, without allowing complementary modes of description, that is a reasonable conclusion. Once one grasps the significance of this apparent contradiction one can only agree with Monod that this is the 'central problem of biology'.

Although the structure-function dualism is one expression of complementarity in living systems, I believe a more explicit epistemological dualism is observable directly as the classical genotype-phenotype distinction where the genetic DNA is to be literally interpreted as a symbolic representation of the aspects of physical structure that are responsible for cell function. In other words, the genotype-phenotype distinction can be understood quite literally as a symbol-matter (or subject-object) dualism. It is just as irreducible to a single mode of description at this primitive level as is the mind-matter dualism at the highest cognitive level. An explanation requires two modes.

In the subjective mode we describe the DNA as a linear, discrete string of symbols that are read and interpreted by the cell's protein synthesizing constraints. This rule-governed reading process, like any other linguistic or computational symbolic process, is rate-independent in the sense that the meaning of the DNA is relatively unaffected by how fast it is read out. Furthermore, the structures of the symbol vehicles are as arbitrary as in any other language and there is no meaning that can be derived from these structures outside the context of the cell that reads and transduces the symbols into functioning enzymes and structural proteins.

In the objective mode we describe the DNA and its interactions within the cell as a three-dimensional, rate-dependent, dynamical system that obeys the causal laws of physics and chemistry. In this mode the DNA and the enzymes have a complex dynamics but no meaning or function that is derivable from this mode. It is this mode that leads reductionists to claim that life is nothing but ordinary physics, which indeed it is as long as one is not willing to consider the subjective problems of measurements and descriptions, goals and policies. What the principle of complementarity says is that using only this one objective mode of description not even physics is reducible to this mode!

8.6 Complementarity Applied to Social Systems

I have developed the concept of complementarity at the simplest levels where it is best known and most certainly needed for adequate explanatory models. How do these concepts apply to the higher levels of social organization? What corresponds to the distinction between laws and rules for organizations that are removed by billions of years of evolution from direct dependence on the laws of physics and whose descriptive modes often occupy such a large fraction of their time?

The basic epistemological need for complementarity is based on the subject-object distinction, and even though social systems observations are many hierarchical levels away from the measurement of electrons, there is still the need to distinguish the dynamical ‘laws’ of social organizations from the descriptive rules that we call goals, plans or policies. Here, the concept of law does not have the degree of inexorability or universality as does physical law, but relative to the goals of the individual it has similar effects. Certainly, one characteristic of social dynamics is its rate-dependence as opposed to the rate-independence of our descriptive social plans and policies. It is important to remember that rate-independence is not the same as time-independence. Just as computer programs may change in time so may governmental policies change in time; however, in both cases, the significance of the result may depend on the order but not the rates of execution. On the other hand, the dynamics of social interactions cannot be turned off or made to wait while policies are being decided. The rate of social change is therefore functionally related to the force and reaction laws of the system. Plans and policies may be said to act as constraints on social dynamics, just as machines act as constraints on physical law. In neither case, however, can we say that the constraints are reducible to the laws or the laws reducible to the constraints. Any explanatory theory must therefore present complementary representations for both modes of activity.

As I said at the beginning, complementarity is primarily an explanatory or epistemological principle, not a prediction and control principle. It is quite possible to form inductive or statistical models from sufficient data to provide extrapolations into the future (e.g. Meadows et al. 1972). Such models may be represented formally by rate equations, but this does not necessarily imply that there is causal rate-dependence of an explanatory dynamics or that there are no policy constraints tacitly mixed into these models. Mesarovic and Pestel (1974) have empirically recognized the need for both causal dynamics and non-causal policies in their computerization of a world systems model, however the principle of complementarity is not recognized in their approach. Therefore, while formal procedures of data processing may lead to useful predictive models of social behavior, there is very little evidence that a deeper conceptual or explanatory theory of social systems ever originates from these procedures alone (see Ghiselin 1955; Polanyi 1962). Social systems theories, in spite of their complexity, are subject to the same epistemological requirements as simpler biological and physical systems where the subject-object distinction of system and measurement-of-the-system cannot be evaded (see Pattee 1972).

The concepts of measurement in physics, of genetic description in biology and of policy in social science may appear disjoint. However, I use them in a hierarchical sense of inclusion, where policy is the most inclusive term. That is, a genetic description is a rather restricted form of policy and a measurement is a rather restricted form of description. The common features of these concepts are that they ‘stand for’ something else, they are symbols or the subject side of the subject-object duality, they are interpreted or executed by constraints, they are rate-independent, and they are not derivable from nor reducible to natural laws.

8.7 How Are Social Dynamics and Social Policy Distinguished?

A naïve answer to this question is that dynamics is what happens because of the external forces and interactions of society, while policy is what some subgroup or individual intentionally decides to implement as a constraint on these dynamics in order to achieve a desired state or goal. The problem is that laws of social actions are not objectively separable from the network of constraints that have been imposed directly or indirectly by past or present policies. Furthermore, many policies appear as simple reactions to social dynamics where the values and goals are not explicitly defined or even recognized.

Instead of trying to distinguish social dynamics and social policy by inductive laws or the assumption of intent on the part of policy-makers, I have proposed the conceptual test of whether the execution or hypothetical model of the activity is rate-dependent, in which case I call it a dynamics, or rate-independent, in which case I call it a policy. I emphasize ‘conceptual test’ rather than formal test, because as I have said, the data without an underlying model can be interpreted either way. Rate-independence must also be interpreted in the causal sense rather than the phenomenological or behavioral sense. As a somewhat artificial example, consider a game of American football. In this case, each play is a highly rate-dependent social dynamics in which the outcome is exceedingly sensitive to small changes of rate. The huddle, on the other hand, is a social policy activity in which the outcome is not sensitive to rate (excluding ‘delaying the game’ penalty). The policy in this case acts as a constraint on the dynamical behavior; but it is also clear that one team’s choice of policy is based on both its model of the game dynamics and its model of the other team’s policy, i.e. complementary models are a necessary condition for rational action. Although the social dynamics and social policies are artificially distinguished in this example, I am proposing that the rate-dependence of dynamics and the rate-independence of policy is the fundamental conceptual distinction in all complementary modes of explanation.

8.8 Dynamical Instabilities and Control Decisions

Having made this distinction, which creates two formally incompatible modes of description, we must now show how these two modes can relate to each other, and how from this interaction we derive explanations or theories of biological or social behavior. How does a system described by dynamical equations that apparently determine its future states from its past also behave as a controllable system where a rate-independent decision can suddenly redirect its trajectory?

The engineering approach to this question involves the concept of parametric control, which is simply the use of additional quantities used in dynamical equations of motion that are not the variables representing the dynamical system itself. However, in so far as these control parameters can be described by additional equations that are functions of rates, the system as a whole becomes a larger dynamical system with no ability to represent decision-making processes.

In order to justify the need for a complementary mode of description we must look for significant situations where one mode of description fails, and fails in such a fundamental way that it cannot be patched up by adjustments of parameters or functional forms. Such situations are well known in dynamical descriptions. They are called singularities by mathematicians and instabilities by physicists. Poincaré characterizes instabilities as situations where 'unobservable causes have observable effects', and more recently Thom (1975) has elaborated this concept into his catastrophe models of control behavior. One does not find explicit recognition of complementarity in Thom's work since the mathematical significance of his method emphasizes the rate-independent topological theorems. Interpreted physically, Thom's theorems are about asymptotic or equilibrium states; and while his method has been used to model a wide variety of biological, social, economic and political behavior (e.g. Zeeman 1976), I believe its explanatory power as well as its susceptibility to empirical test would be greatly strengthened by explicit recognition of the rate-dependent dynamical laws on an equal footing with the topology, that is, as a principle of complementarity. I would argue, in fact, that the explanatory power of Thom's method arises from his tacit use of an underlying dynamical language, augmented by his attention to precisely those areas of behavior where this traditional dynamical mode fails, i.e. at singularities where the order (topology) of events takes precedence over the rate of events (Pattee 1976).

8.9 Complementarity and the Evolution of Complexity

The unique character of biological and social systems behavior that distinguishes them from non-living systems is their tendency to evolve greater and more significant complexity. This is true from the cell to society; and in terms of the statistical concept of natural selection this theory can be, in part, represented by dynamical models. The complementary rate-independent description of evolution in terms of fitness is still in a rudimentary form, since fitness in its global form is the totality of nature's 'policy'. However, there is clearly a rate-independent hierarchical organization of living systems and this complementary model allows some chance to look at one or two levels of structure and function at a time. My point is that unless the complementarity of structure and function, or pattern and process, are represented by the model we will not have an explanatory theory. As we pointed out at the beginning, the representations of the molecular structure of the genetic DNA and of the symbolic process it represents are expressed in separate languages that are formally incompatible. It is significant that the only acceptable theory of evolution requires both a selection dynamics based on the differential birth and death rates of a population that can only be explained in phenotypic concepts (i.e. selection does not act on the genotype); and also hereditary rules that constrain the phenotype that can only be explained as rate-independent descriptions. By analogy, we may say that population dynamics are constrained by genetic policies, but the dynamics do not directly affect the policies, nor do the policies directly affect the dynamics. It is the indirectness of the influence of structure on function at the biological levels and of policy on behavior at the social levels that make comprehensive prediction so difficult. Even at the simplest levels of biology we do not have the knowledge to relate structure and function. For example, we have no predictive theory of how a change in the primary sequence of amino acids in an enzyme will affect its specificity or catalytic function; even though the explanatory relations between structural genes and enzyme operation are well developed. It is certainly no wonder then that we cannot predict how a genetic change will affect the social behavior of an individual, when in most cases we do not even have an explanatory picture of how genes are expressed at epigenetic or behavioral levels.

8.10 Conclusion

Explanatory theories of biological and social systems are presently limited by the classical paradigm of explanation that requires unified, consistent, objective models. This paradigm is derived from the physical sciences of the last century. An epistemology that explicitly recognizes the individual as an intentional agent in all observation and control processes was forced upon physics only with the formulation of quantum theory, but the generality of this subject-object dualism has never been adequately recognized in the normal thinking of biological, behavioral or social scientists. The greatest difficulty with the concept of complementarity, besides the fact that it is not now a generally acceptable paradigm, is that its formulation and application have not been developed in a broader context than quantum theory where, unfortunately, the epistemological problems of measurement are still not clearly resolved. A second difficulty is that complementarity is an explanatory principle having to do more with the inner consistency of models of observational situations rather than simple simulations that predict results. It is now only an epistemological principle, not a practical engineering principle. Its acceptance in quantum mechanics only came about because of the failure of every other interpretation. This may also be the only hope of its incorporation into biological and social theories.

References

Bohm, D. (1971). On Bohr's views concerning the quantum theory. In T. Bastin (Ed.), *Quantum theory and beyond* (pp. 33–40). Cambridge: Cambridge University Press.

Bohr, N. (1928). The 1927 Como lecture reprinted in *Nature 121*, 580.

Bohr, N. (1937). Biology and atomic physics. In *Atomic physics and human knowledge* (pp. 13–22). New York: John Wiley.

Bohr, N. (1938). Natural philosophy and human cultures. In *Atomic physics and human knowledge* (pp. 23–31). New York: John Wiley.

Bohr, N. (1949). In P. A. Schilpp (Ed.), *Albert Einstein: Philosopher-scientist* (pp. 201–241). Evanston, IL: The Library of Living Philosophers.

Ghiselin, B. (1955). *The creative process*. New York: New American Library (Mentor).

Haldane, J.B.S. (1963). Discussion at conference on the origin of prebiological systems, Wakulla Springs, FL, October 1963.

Jammer, M. (1974). *The philosophy of quantum mechanics*. New York: John Wiley.

Meadows, D. H., Meadows, D. L., Randers, J., & Behrens, W. W. (1972). *The limits to growth*. New York: New American Library (Signet).

Mesarovic, M., & Pestel, E. (1974). *Mankind at the turning point*. New York: New American Library (Signet).

Monod, J. (1971). *Chance and necessity* (p. 21). New York: Alfred Knopf.

Pattee, H. H. (1972). Laws, constraints, symbols, and languages. In C. Waddington (Ed.), *Towards a theoretical biology* (Vol. 4, pp. 248–258). Edinburgh: Edinburgh University Press.

Pattee, H. H. (1976). The role of instabilities in the evolution of control hierarchies. In T. Burns & W. Buckley (Eds.), *Power and control—Social structures and their transformations* (pp. 171–184). London: Sage Publications.

Polanyi, M. (1962). *Personal knowledge*. London: Routledge and Kegan Paul.

Thom, R. (1975). *Structural stability and morphogenesis*. Reading: W. A. Benjamin.

von Weizsäcker, C. F. (1949). Zur Deutung der Quantenmechanik. *Zeitschrift für Physik, 118*, 489.

Zeeman, E. C. (1976). Catastrophe theory. *Scientific American, 234*, 65.

9 Clues from Molecular Symbol Systems

Abstract Natural selection operates on living systems through their function and behavior. The biological structures constraining this behavior always involve fortuitous elements, or frozen accidents, as well as essential principles. In order to distinguish the accidents from the principles we must refer to some theory of living systems. Similarly, in order to distinguish which biological constraints on linguistic form are fortuitous and which are fundamental, we must refer to some theory of symbolic systems. A theory of symbols must address the process that relates the symbol vehicle to its referent or meaning. At the level of natural language we have many facts, but still have great difficulty incorporating them in a theory of language. However, at the level of the gene the relation of symbol structures to their referent function is better understood. A careful look at this elementary symbol system may provide some clues to basic principles of language at higher levels. In particular, we consider the articulation of the discrete, rate-independent, linear symbol strings, which generate continuous, rate-dependent, three-dimensional functions through the folding transformation. We suggest that this complementary interaction of constraints and laws involves general principles that are elaborated in higher linguistic forms.

9.1 Background

I am going to describe a biological symbol system that has many of the properties that linguists often use to distinguish human language from other forms of communication. My purpose is not to make any claims about whether this system is or is

---

Reprinted from *Signed and Spoken Language: Biological Constraints on Linguistic Form,* U. Bellugi and M. Studdert-Kennedy, Eds. Dahlem Konferenzen 1980. Weinheim: Verlag Chemie GmbHpp., pp. 261–274.

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7,
DOI 10.1007/978-94-007-5161-3_10,

not a language. Rather, I want to show some of the basic physical and logical requirements that are necessary to support this symbol system and to suggest that these requirements at this most elementary level may provide some useful clues to the universals of language at the highest levels.

But more than that, I shall try to present the rudiments of a theory of symbol systems that may also provide some clues for a theory of language. By a theory I mean a model that tells you what is important, as distinguished from what may be true but unimportant. Unfortunately many structures in living systems can be widespread or even universal without being important to a theory of life. The same is true for language universals as many linguists have pointed out (e.g., Hockett 1966).

The system I am talking about is the genetic system in cells. Now many linguists as well as philosophers of language may feel that DNA, and all that, is at best an overworked metaphor and at worst a collection of macromolecules looked at by an animist, but there are notable exceptions. Roman Jakobson (1970) wrote, " … one could venture the legitimate question whether the isomorphism exhibited by these two different codes, genetic and verbal, results from a mere convergence induced by similar needs, or whether, perhaps, the foundations of the overt linguistic patterns superimposed upon molecular communication have been modeled directly upon its structural principles." Again, I do not wish to argue whether genetic expressions are to be called a language or not. The only useful approach is to study explicitly the structural principles that support this primeval symbol system. We must then have a theory of symbols systems before we can discuss the importance of these principles in other domains of communication and expression.

9.2 The Physics of Symbols

A symbol is defined as something that stands for something else by reason of a relation. At the molecular level an obvious question is what supports or executes the relation between the symbol and what it stands for. In physics we cannot invoke incorporeal relations such as definitions, associations, or connotations. We are stuck with the global laws of nature and the local constraints of physical structures to establish all relations. A symbolic relation, however, is clearly not the same as an interaction or collision. We do not say that the moon is a symbol for the earth because it has a gravitational relation with the earth, nor do we say that a nucleotide is a symbol for polynucleotides simply because it is related by a chemical bond to other nucleotides. What makes us aware of the polynucleotide DNA as a symbol is the existence of a specific and separate set of enzymes called the synthetases along with the transfer RNAs and ribosomes that execute the code relation. In other words, any "stands for" relation must have an explicit embodiment in the form of constraints that are not an inherent property of either the symbol vehicle or the referent. The question is what are the important conditions for these symbolic constraints and what are merely fortuitous or arbitrary frozen accidents.

For example, the laws of chemistry allow left- and right-handed polynucleotide helices as well as left- and right-handed amino acids with no preference between them. We find it is a universal fact of life that only one-handedness is used. However, no one believes this universal is important for a theory of life. By contrast, the fact that the synthetases and transfer RNAs have their own sequences described in the genetic DNA is believed to be of fundamental importance for a theory of the origin and evolution of life (e.g., Eigen and Schuster 1977; Pattee 1977). The point is that without some theoretical framework, there is no way to interpret observations. Superficially, a theory generates and accepts certain observational data strings very much like a grammar generates and accepts certain word strings. What we are missing is at a deeper level where the symbol strings generate meanings and comprehension.

This relation between symbol vehicle and its referent in ordinary language has proved to be incredibly difficult to pin down. In philosophy one cannot discuss this relation without becoming enmeshed in epistemological assumptions that appear inscrutable. So let us look at the simplest case we can find of a symbolreferent relation and see what actually takes place. Even at this simplest level there are many complications. I shall only outline what I believe to be the essential features according to my rudimentary theory of symbolic activity.

Symbol systems require, almost by definition, a rather small set of symbol vehicles. At least the set must be closed; otherwise the structures that execute the referent relations would need to be multiplied indefinitely. This also implies discrete symbols or symbol classes. In DNA there are only four basic symbols, and it is plausible that at the origin of life there were only two. However, like natural languages, DNA has a multiplicity of patterning forming lexical and syntactic hierarchies, and I would claim further, a semantic hierarchy. Omitting as many details as possible, I come directly to the relation between a structural gene and its referent, say, a particular enzyme molecule. How is this relation executed, and what are the essential properties of the relation?

First, it is clear that many physical and logical aspects of a gene-enzyme relation are arbitrary frozen accidents—a kind of convention. The genetic code itself appears to have no basis in the shape, size, or chemistry of its components. Nor do the translating enzymes that help execute the code appear as necessarily unique structures. There is no logical or physical reason known why we could not have the same type of life using an entirely different code and even different primary sequences for enzymes. The only logically essential feature is the coherent function of the system, i.e., the code must be complete, the symbols and their translating molecules must form a closed set, the translating molecules must be describable by the genetic symbols, and the end result, that is, what the symbols "stand for" must appear in their functional form.

I want to emphasize that what a symbol stands for cannot end with just another symbol. The referent of a symbol is an action or constraint that actually functions in the dynamical, real-time sense. Here is where any formal language theory loses contact with real languages. A formal language only generates symbol strings from other symbol strings according to the rules of a grammar, and a fundamental restriction on these rules is that they are rate-independent. Formal languages are therefore excluded from the world of real molecules as well as real behaviors that occur as rate-dependent interactions (e.g., von Neumann 1966). An important property of the gene-enzyme or symbol-referent relation, then, is that the symbols exist in a rate-independent context whereas their referents function in a rate-dependent context. In other words, the mechanism that executes the relation between symbol and referent must use rate-independent articulations (usually one-dimensional discrete strings) to constrain rate-dependent functions or action (usually threedimensional, continuous behavior).

What types of constraints do this, and what are their physical characteristics? In physics we usually like to think of either stable, time-independent structures, like molecules, crystals or tables, or moving objects that can be described by rate-dependent equations like particles, waves, or fluxes. The constraints that translate symbols must be thought of as a combination of stable but movable (i.e., flexible) structures that do not move according to rate-dependent laws. In fact most machines are constructed from such constraints. They are called non-holonomic or nonintegrable constraints since there is no rate expression to integrate (Pattee 1977). However, the artificial machines that we construct from non-integrable ratchets, bearings, and gears, or in the case of computers, from switches, gates, and clocks, are not constructed on the same principles as the non-integrable biological constraints of which the enzymes are the most basic example. In fact the nature of the symbol-referent relation is quite different in natural and artificial symbol systems, and this is one reason we find it so difficult to model language, or to find an adequate theory of language using the computer or other manifestations of the machine paradigm of explanation that dominate the classical sciences.

9.3 The Symbol-Folding Transformation

What is distinctive about the symbol-function relation in the gene-enzyme transformation that is not found in artificial or formal symbol-function relations? Let us look at what happens to the DNA symbols as they are transformed into their referent, i.e., the functioning enzyme. What, exactly, do these genetic instructions do to achieve their ultimate meaning or function? First, they are literally translated from the nucleotide strings of symbols to the amino acid strings. The relation between these two strings is called the genetic code, but it does not tell the whole story by any means! Once the one-dimensional amino acid string is sequentially synthesized the genetic symbols cease any further instructional activity. The information processing has been completed, and now this symbolic sequential order serves only as a remarkable non-integrable constraint harnessing the universal forces of nature to produce a new, three-dimensional folded structure that becomes more than a structure. It is now a precisely functioning dynamical machine that can speed up the rate of a specific reaction by factors of 10^6–10^{12}.

Artificial machines are not constructed so cleverly. It is as if we could design any machine so that it could be assembled simply by hooking the parts together in a chain, and then have the chain spontaneously form itself into a functioning mechanism. In other words, the genetic symbols are not related to their referent action in any detailed or explicit form, but only through an implicit harnessing of natural laws and structures which need no instructions. In fact the amount of information in the genetic symbol string is only a very small fraction of the information that would be necessary for a completely formal and explicit specification of the structure of the enzyme.

There are certainly no symbols in the gene for the three-dimensional coordinates of each amino acid residue, let alone for each atom in the enzyme. Life would hardly be possible if such symbolic detail were necessary, since the mass of each gene would far exceed the mass of the cell it could describe. But the folding transformation has even more significance than its demonstration of the implicit nature of the relation of symbols to their referents. The relation of the individual genetic symbols to the function of an enzyme is not localized in some corresponding step or component of the function. Changes in genetic symbols do not effect proportional changes in the function. In some cases, numerous, discrete changes in the gene may produce only small and continuous modulation in the enzyme function, while in other cases a single base change in the gene may produce nonsense or even a lethal mutation.

9.4 Is the Folding Transformation Universal?

There is little doubt that the folding transformation is of utmost significance for the origin and evolution of life, but does it give us a clue to theories of natural language? There are, of course, theories of language that are quite consistent with the behavior of the folding relation. Merleau-Ponty (1964), elaborating on Saussure's theory that signs are never explicit in their meaning, says, "(…) expression is not the adjustment of an element of discourse to each element of meaning (…). To speak is not to put a word under each thought; if it were, nothing would ever be said." The Cartesian idea that the grammatical rules governing the strings of symbols were only a surface structure that are inadequate to derive the meaning of the symbols is also consistent with a folding transformation (e.g., Chomsky 1972).

Of course the nature of the deep structure is the central issue. The laws of physics which are responsible for the deep structure of the primeval genetic symbol-referent relation have certainly been augmented by higher levels of genetic frameworks that may perform a type of folding transformation on higher level symbol strings to obtain what we call their meaning. Nevertheless, any transformation from linear, discrete, sequential, rate-independent symbol strings to the three-dimensional, continuous, highly parallel, rate-dependent function should be carefully considered, at least as a conceptual basis for a theory of linguistic competence. There is no reason to expect that at the level of the brain the structures executing this type of transformation are like synthetases and ribosomes. On the other hand, if the principle of harnessing the deep structures by a folding transformation using the surface structures of symbol strings has worked for evolution, there is reason to expect that it would also work for learning processes. How could this folding process at the molecular level have significance at higher levels of organization? Does it provide clues to learning and language? In other words, is folding generalizable? I believe it is. What happens in the construction of a cell is that a very explicit symbol sequence in the genetic DNA produces other explicit strings of polypeptides. Then, quite abruptly, we lose track of the explicit symbols.

The polypeptides become autonomous—to form enzymes, they fold themselves into precise machines, and to form structural proteins they self-assemble into fibrils, membranes, and muscles. This self-assembly process is not an explicit symbol processing, but a harnessing of the laws by these symbol-dependent constraints or boundary conditions (Pattee 1972; Polanyi 1968). At this second level of self-assembly, after the folding of single chains, we now see each protein molecule constraining its neighbors by its unique configurations to form higher structural levels with corresponding higher level functions. Thus all these "foldings" lead to a hierarchy of structures and functions, each harnessing the structures and functions of the lower level. This ontogenetic hierarchy of folding, self-assembly, morphogenesis, epigenesis, and development are all recognized as being ultimately "under genetic control," although the mechanisms at higher levels remain largely unknown.

The relation of the verbal symbol strings to meanings in natural language is quite similar. We begin with strings of words which have explicit transformational equivalents. In other words we can, up to a point, translate symbol strings into different symbol strings in the same language or into a different language while preserving the meaning of the original string. But this "information processing" is a surface transformation. It only goes so far towards generating meanings. At some point we recognize that our explicit strings have somehow generated a meaning and that this meaning is now largely autonomous; that is, we now know the meaning without needing the direct support of the detailed explicit constraints of the symbol strings. Furthermore, this meaning now becomes a constraint on higher level meanings. Perhaps it is only metaphor to say that these word strings are folded under the rules of the cognitive structures in the brain to generate the meaning, but I believe the metaphor is suggestive. We know that meanings of words and sentences have the same hierarchical openness as the functions of amino acids and proteins. At each level of organization there appears a function or meaning in which a given symbol string plays a role appropriate to that level. Each string has many functions or meanings depending on the level of activity being constrained. We do not so much construct meanings by building with words as we use words to constrain the natural images (i.e., the deep structures) of the brain. This is why so few genetic symbols can generate such elaborate structure and function, and "how so little information controls so much behavior," in Gregory's (1969) words.

We do not always have to build up explicit meanings from scratch, bit by bit, since the brain, like the cell, is already endowed by evolution with hierarchies of functions which only need to be constrained or harnessed to generate the meaning coded in the symbol strings. Just as the highest levels of biological organization are recognized as "under genetic control" so our most subtle and abstract meanings appear to be "under linguistic control," but in both cases the actual symbolic input appears totally inadequate for the job of control by any classical machine paradigm. It was the recognition of this inadequacy that prompted Waddington in biology to invent the epigenetic landscape, and the Port Royale philosophers in linguistics to invent the deep structures "present in the mind" that Chomsky associates with linguistic competence. Both theories are still vague in their mechanisms, but both in effect postulate a rich body of more or less tacit and general powers that are able to execute specific functions or actions only under the constraints of the explicit strings of genetic or verbal symbols (cf. Polanyi 1958). Therefore, in so far as the generation of meanings from symbol strings involves the use of the string to form hierarchies of structural constraints on already existing dynamical organizations, the concept of folding is a plausible model of considerable generality. The practical value of the molecular instance of folding is that unlike other theories, such as epigenesis or competence, we have some observable clues as to what processes are important for a symbol-function transformation.

What does folding accomplish? What is actually being transformed at the molecular level? As we indicated earlier, the particular molecules used in the nucleic acids and proteins, as well as the genetic code that translates one molecular alphabet to the other, appear to be largely arbitrary. In other words, the shapes and activities of the alphabetical symbol vehicles are not iconically or mimetically related to any elements of their function. Nor do the structures of the linear sequences of these symbol vehicles come any closer to their function simply by being translated from one molecular alphabet to another. At this information processing or surface level of grammar, all the lexical and syntactical rules apply sequential: and locally, i.e., are mostly context-free, and the behavior of the strings have no rate-dependence. However upon folding, the string acquires a binding site for a substrate molecule and a rate-dependent catalytic activity. The structure of the folded string has both a shape and an activity, neither of which could be called arbitrary. In fact, the folded enzyme can be called both iconic in its binding site and mimetic in its catalytic activity. The folding process as well as the substrate binding and catalysis are no longer local, sequential, or context-free, but highly parallel and global. Yet physically the folded enzyme contains the same molecules and obeys the same laws as the string of symbol vehicles. What has happened is that those aspects of structure that are essential for the activities as symbol vehicles have been suppressed, while other aspects of the symbol vehicle structure, that are excluded from explicit symbolic activity, have been brought into play. Furthermore, these two aspects appear as complementary in the sense that one aspect cannot function in both roles simultaneously (Pattee 1979). A structure cannot be both arbitrary and iconic, or rate-independent and rate-dependent, or discrete and continuous at the same time. However, the same structure can possess any of these aspects depending on its context, i.e., the level of organization where the particular aspect is brought into play. For example, an amino acid molecule at the symbol vehicle level of organization depends on covalent reactions of the so-called backbone chain, and is independent of weak side-chain interactions. On the other hand, the folding depends essentially on the weak side-chain interactions while the covalent bonds act only as passive constraints.

9.5 Conclusion

From the characteristics of this primitive symbol system we find several suggestive clues to a general theory of language. First, we have more or less direct physical evidence that the constraints acting on the linear, discrete, rate-independent symbol strings are inadequate for any process of interpretation. We may compare this symbol string grammar to the ordinary surface structure of our natural language grammar. It is significant that at both the genetic and verbal levels this surface grammar is easily describable within its own symbol system.

We also associate a high degree of convention or arbitrariness with this level of symbolic activity. By contrast, the folding transformation brings into play forces and constraints that are not within the symbol system itself. With respect to genetic descriptions we could call these forces ineffable. Also, we see evidence of this folding transformation suppressing the arbitrary aspects of symbol vehicle structure and generating both iconic and mimetic behavior from complementary constraints of the symbol vehicle strings. We may compare the role of these complementary constraints to the postulated deep structure of the brain. Perhaps, as many philosophers have suggested, some of these deeper constraints will remain ineffable with respect to the language they serve to interpret. However, even if this should be the case, it would not mean that an adequate theory of language is impossible. In fact, it is my principle conclusion that general principles of symbol systems can be found that are as universal as the principles we call natural laws. There are many levels of symbolic activity between the gene and human language where I believe these principles can be studied.

References

Chomsky, N. (1972). *Language and mind*. New York: Harcourt Brace Jovanovich.

Eigen, M., & Schuster, P. (1977). The hypercycle. A principle of self-organization. *Naturwissenschaften, 64*, 541–565.

Gregory, R. L. (1969). On how so little information controls so much behavior. In C. H. Waddington (Ed.), *Towards a theoretical biology* (2nd ed., pp. 236–274). Edinburgh: University of Edinburgh Press.

Hockett, C. F. (1966). The problem of universals in language. In J. H. Greenburg (Ed.), *Universals of language* (pp. 1–29). Cambridge, MA: MIT Press.

Jakobson, R. (1970). *Main trends of research in the social and human sciences* (pp. 437–440). The Hague/Paris: Mouton/UNESCO.

Merleau-Ponty, M. (1964). Indirect language and the voices of silence. In M. Merleau-Ponty (Ed.), *Signs*. Chicago: Northwestern University Press.

Pattee, H. (1972). Laws and constraints, symbols and languages. In C. H. Waddington (Ed.), *Towards a theoretical biology* (pp. 248–258). Edinburgh: University of Edinburgh Press.

Pattee, H. (1977). Dynamic and linguistic modes of complex systems. *International Journal of General Systems, 2*, 259–266.

Pattee, H. (1979). Complementarity vs. reduction as explanation of biological complexity. *American Journal of Physiology, 5*, R241–R246.

Polanyi, M. (1958). *The logic of personal knowledge*. New York: Harper & Row.

Polanyi, M. (1968). Life's irreducible structure. *Science, 160*, 1308–1312.

von Neumann, J. (1966). *The theory of self-reproducing automata* (completed and edited by A.W. Burks) (Chapter 5). Urbana: University of Illinois Press.

10 Cell Psychology: An Evolutionary Approach to the Symbol-Matter Problem

Abstract I describe the simplest living organism, the cell, as a symbol-matter system—an observable case of how a natural representation using a word processing format constrains the real-time behavior of a material organism. My purpose is not to model cognition or language at the brain level, which cells have achieved only after some three billion years of evolution. Rather, I use this primeval embodiment of a symbol-matter system as an exercise in mental hygiene for cognitive scientists to test, and hopefully improve, the clarity of their fundamental explanatory concepts. I focus on the assumptions of information processing and direct perception approaches as specific examples of inadequate theories for understanding even simple symbol-matter systems. Based on the genetic organization of cells, I propose a semantic closure principle that requires both the physical constraints of direct perception and the syntactic constraints of information processing for explanation of symbol-matter systems. This exercise is motivated by the belief that if we expect to get anywhere with the mind-body problem at the brain level, then our concepts must at least be adequate in scope and precision to explain the symbol-matter relation in single cells where it all started.

10.1 Background

10.1.1 What Is the Problem?

Adequate explanation in any cognitive science must at some stage address the matter-mind problem, that is, the problem of symbol reference or how the world and our images of the world are coordinated. However, this problem has proved so difficult

---

Reprinted from *Cognition and Brain Theory*, 5(4), 325–341, 1982.

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7,
DOI 10.1007/978-94-007-5161-3_11,

over the centuries, that not only is there no apparent convergence toward some solution, but there is no consensus on what is required of any solution. So cognitive scientists have a secondary problem of choosing what types of investigation they can do if they can't attack the matter-mind problem directly. The most popular choice today is the study of artificial symbol systems. This approach assumes a computational or formal basis for the mind, leaving the matter side of the problem alone as conceptually unnecessary or as realizable in principle by more or less arbitrary material transducers, like robots. In contrast to this computational or information processing approach to cognition there are the ecological realists or direct perceptionists who study the material constraints of the perceiver-world system, leaving the symbolic side of the problem alone as either artificial or non-existent. Of course, in between these largely disjoint approaches are the more or less empirically oriented scientists who study the brain itself, believing that nervous tissue is the place where matter and mind must either connect up, hopefully in some observable way, or perhaps where the distinction will just fade away.

Each of these areas has its own characteristic secondary problems, of course, but all of them have two very basic uncertainties in common. First, since none of them attacks, or even defines, the mind-matter problem directly, there is little assurance that any of these approaches is converging toward an answer to that basic problem. Indeed, the information processors and direct perceptionists have more or less explicitly avoided the problem by the way they have chosen their commitments. The second basic problem is that each has chosen independent primitive concepts, and has therefore generated more or less unrelated styles and dialects which are growing in formality and complexity to the point where even now it is difficult for outsiders to follow. Should the present trend continue, we will end up at best with a collection of pure but incoherent models. Of course there is an additional pessimistic possibility that the human brain is in fact too knotty a problem to be unraveled by simple models; that to behave like a brain we need a model, in whatever language, that is as complex as the brain. This argument may be used to support computer simulation as the only model that can imitate such great complexity. Yet, even if a computer copy of a brain were to be achieved, it is not clear how such an isomorphism, in itself, would provide any explanation of the matter-mind problem (Davidson 1981).

In the meantime, even general arguments about the foundations of cognitive science do not sound convergent, for example, consider Minsky's (1980) attitude in his criticisms of Searle, where Minsky concludes:

> Yet a feeling remains that there is something deeply wrong with all such discussions (as this one) of other minds; nothing ever seems to get settled in them. From the finest minds, on all sides, emerge thoughts and methods of low quality and little power. Surely this stems from a burden of traditional ideas inadequate to this tremendously difficult enterprise. Even our logic may be suspect [p. 440].

I was reminded of Newell's (1973) earlier assessment of the possibility of resolving the fundamental oppositions in psychology:

> As I examine the fate of our oppositions … it seems to me that clarity is never achieved. Matters simply become muddier and muddier as we go down through time. Thus, far from providing the rungs of a ladder by which psychology gradually climbs to clarity, this form

> of conceptual structure leads rather to an ever increasing pile of issues, which we weary of or become diverted from, but never really settle [p. 289].

Of course both Minsky and Newell then restate their commitment to still more information processing as the best hope for finally "settling" these problems. Minsky argues that we must rely on the "empirical power" of our methods to pursue these problems, and that computationalism has shown this power. The ecological realists would argue that this is only artificial empiricism (e.g., Carello et al. 1982). The realists on the other hand have as yet no empirically testable working models of their own.

10.2 Simple Psychologies

Where can the cognitive sciences expect to find more promising foundations for symbol-matter theory? I see no course but to look for the first "rungs of the ladder." We need simpler embodiments of *natural* matter-symbol systems with both empirical power and conceptual generality. Why should we work only with the ultimate functional complexity of brains, or the ultimate artificiality of computers, or the ultimate meanings of philosophical discourse? As a trial first rung I suggest trying to adapt our fundamental concepts of cognitive science to the basic symbol-matter problems of biology, and even physics, where a few rungs have already been secured.

There are significant properties of natural, symbol-matter systems that show up empirically in biology and conceptually in physics that are totally evaded by present computer and direct perception models. Also, I suggest explicit recognition of evolutionary principles in our choice of basic concepts. Brains have only gradually evolved from cells. Both the material structures and functions have coevolved. The types of behavior we talk about in brains have also evolved. This means that high level concepts such as intentions, meanings, thoughts, and so on, which we associate only with minds, must have had evolutionary precursors in a more or less gradual sequence. The problem is that we do not have a clear concept of what the simplest "intention," "meaning," or "thought" might look like. This is because psychology has traditionally been defined by only highly evolved "mental" activity, so that even though we study brains at the cellular or even molecular levels, there is the tacit belief that no real psychology can exist at a simple level. Human physiology was also first defined as the study of gross organs and body fluids, but gradually these concepts were generalized and refined by the study of simpler organisms until today we find the foundations of human physiology in cell physiology. This does not mean that cells explain or exhibit all higher level processes. Cells do not have feet or ears, but they have motility and irritability, which are basic functions of feet and ears. Physiologists have no problem accepting computer simulations of complex physiological systems; but at the same time no physiologist would claim basic understanding without knowing the concepts and facts of cell physiology.

10.2.1 Cell Psychology

I suggest that brain-level psychologies are not likely to converge until we have some agreement on the foundations of cell psychology. Obviously I am not suggesting that cells have minds any more than cells have feet, but cells are certainly symbol-matter systems which hopefully can be more clearly understood and modeled. If we cannot agree on a set of concepts and a language that describes the symbol-matter behavior in cells, then our disagreements at the matter-mind level are certainly never going to be settled. Of course, even if we can agree on the cellular level of description there is no guarantee that we can agree at the brain level. However, that is a later problem. My own feeling is that finding an acceptable set of concepts will be difficult even at the cellular level, for reasons which will become apparent. What I do expect is that by observing more carefully the natural symbol-matter relationships in cells, we will see the inadequacies of information-processing and direct perception as explanatory theories.

Before I outline the molecular semiotics and dynamics of cells, let me emphasize that I am not claiming that cells (matter and all) cannot be simulated in detail by a computer, nor that cells (symbols and all) cannot be described in detail by physical principles. If you value such detailed exercises, fine. In fact I am a firm believer in the ability of computers to simulate anything at all, except the ineffable, as well as in the universal applicability of physical laws to all behavior, living and nonliving. This is not an issue of formal or physical reductionism, but an issue of explanatory meaning, or the value of models. The problem and the attraction of physics and computation as bases for models is that they are *both* universal, but complementary, modes of description. Physical laws are universal and inexorable; everything must obey physical laws, we assume, even if our *descriptions* of these laws always have some inaccuracies. Computers are universal and conventional; everything can be described by a computer convention if it can be described by any other convention (i.e., a language). The issue is whether the description of symbol-matter systems by either mode alone is a satisfactory explanation. The cases at point: Do descriptions of symbol vehicles taken only as matter following laws of motion, have explanatory value; or do descriptions of coordinated dynamics taken only as programs, have explanatory value? This is one issue that causes a problem in brain models, and it is one issue that will cause a problem in cell psychology. We need some clarification and agreement on how to answer symbol-matter questions about cells. Is it good cell psychology to say that the gene is a "representation" of the enzyme? Does it represent the enzyme's structure or function or both? Is the enzyme's recognition and catalysis of the substrate dependent on a "direct perception" process or a "formal inductive" process? Does the action of the enzyme require information at all? Is the reading frame of the DNA (the codon) genetically penetrable (cognitively penetrable) or is it part of the fixed architecture of the cell?

If you think the answers to this type of question are obvious, try them on your friends. Bear in mind that any answer must address at least two aspects of the question: (1) do we agree on what the question means, in other words, can we define the words; and (2) is that how it works, meaning, do the facts of molecular biology confirm it. These are both difficult problems, but for the cell at least we know many details of "how it works." The brain, on the other hand, is enormously more complex, and we are nowhere near knowing "how it works."

10.2.2 Are Cells Too Simple?

There are at least three variants of the objection that cells are too simple to be of interest to psychologists. One feeling is that cells are simple enough to be adequately described by chemical language and that if we do speak of the genetic "representation" and substrate "recognition" this is only metaphorical. Another feeling is that whatever information processing the cell does, it cannot be essentially different from what a simple computer can do, and since the physical realization is not important, the cell gives us no more clues to informational behavior than computer models do. A third commonsense feeling is that "obviously" cells are just not complicated enough to have behavior like brains.

The first objection could only be felt by someone who has not been keeping up with the molecular biology literature, which shows that traditional biochemical language is *not* adequate to explain the string-processing activities of the cell. The question of being "metaphorical" is precisely one of the issues in finding a conceptually powerful language. Good metaphors very often guide the best explanations. I will show that the second objection that cells are like computers is incorrect in precisely those aspects of the natural symbol-matter relation that are essential for evolutionary behavior. It is these differences that I hope will give us clues on how to improve our concepts of symbol-matter interactions. The third objection that cells do not behave like brains is quite correct, but not relevant to my argument. My argument is that: (1) we do not agree on what we mean by the symbol-matter problem; and (2) we have no clear empirical tests of our concepts of symbol-matter interaction at the brain level; therefore (3) we should first test our basic concepts at the cellular level where we know more about how it works.

10.3 The Facts of Life

10.3.1 Only Symbol-Matter Systems Evolve

Biology has never been strong on theory. Even when a good theory is found, such as von Neumann's (1966) theory of self-reproducing automata, it is ignored by most biologists. The so-called germ plasm theory of Weismann is more like a fact of life than a theory, without the logic of von Neumann. Evolution is the most general biological theory, although it is anomalous by physical standards, since it cannot predict any detail, and has no universally accepted formal structure. Furthermore, many of its search and selection strategies are still disputed or unknown. However, the point I want to emphasize is that evolution and cell replication, whether we call them fact or theory, have an absolutely fundamental requirement in common—they both need the articulation of symbol and matter, or in the language of biology, the articulation of genotype and phenotype, to succeed. As far as I know, there is no totally convincing explanation of why this is necessarily the case. Von Neumann (1966) speculated that a purely dynamical, direct replicating automaton without a symbolic self-description would not be likely to successfully propagate its physical "mutations" and hence would not evolve. Of course this is not a formalizable conjecture since its truth must depend on the physical details of the automaton's embodiment. I have argued (Pattee 1968) that a symbolic representation of a hereditary process is a strategy to overcome the problem of unreliability of all dissipative constraints (i.e., rule-executing dynamical machines). However, for this discussion I will simply take the existence of the symbol-matter articulation as an *observed* fact of life. To avoid either side of this link is to miss the link. It is significant that when von Neumann 1966 later chose to pursue the pure logic of self-description he was well-aware that he was cutting the link to any empirical symbol-matter questions:

> By axiomatizing automata in this manner, one has thrown half of the problem out the window and it may be the more important half. … One does not ask the most intriguing, exciting and important question of why the molecules or aggregates which in nature really occur in these parts are the sort of things they are, why they are essentially very large molecules in some cases but very large aggregations in other cases [p. 77].

By contrast, the information processors have apparently thrown the material half of the problem out the window, not with curiosity, or even caution, but with a commitment never to look at a molecule as a problem again. The direct perceptionists for their part have apparently thrown the symbolic half of the problem out the window with the commitment never to look at a bit string as an answer again.

It is easy to understand why symbol-matter problems "never get settled" under these two commitments. It is not so easy to understand why so much effort is channeled into defending the exclusivity of two such commitments, both of which evade the symbol-matter problem. One reason, I suspect, is that both sides are so *good* at evading the problem. As I said before, this is because either symbol systems or constrained physical dynamics can, according to their own principles, be used to simulate anything.

What I believe the facts of life show is that information processing and direct perception have necessarily coevolved in a much more intricate and efficient articulation than is imaginable under the present commitments of computationalists and ecological realists. In fact, it is precisely this natural symbol-matter articulation that makes life distinct from non-living physical systems, including our present type of computer. I make no claims about the question of the future creation of artificial life or even artificial mind. If I were required to make a guess, I would bet "yes" in principle, but "no" in practice. All I am arguing is that these two schools of thought are not on the right track for explanatory models or even simulations of simple, natural symbol-matter systems. My first argument is based on experimental evidence from molecular biology; my second argument depends on the conceptual structure of physics.

10.3.2 Molecular Information Processing

One well-established fact of life is that the genetic mechanisms of cells use string processing. Genes are strings of only four characters (nucleotides) that are read sequentially, transcribed sequentially into other strings (messenger RNA), and translated sequentially into strings of a larger character set (amino acids). The gene strings can be described as a formal language with a set of variables (structural genes, operators, promoters, terminators, etc.) perhaps comparable in size to the syntactic classes of natural language, and a phrase-structure grammar (Waters 1982) that looks like a simple programming language grammar. The genetic code which maps the codons to the amino acids is very likely arbitrary to some degree, although the evidence is very sketchy (Hoffman 1975). In any event, there appears to be no physical or logical reason why the same phenotype could not in principle be produced by gene strings in a very large number of different genetic languages. Our present terrestrial genetic language appears to be largely a frozen accident, but we have not verified this by observation of other forms of life. Therefore, if we do not ask "the most intriguing questions" about the origin, evolution, or meaning of genes, they can be described literally as strings in a formal information-processing symbol system. Whether or not the basic operations are sufficient to form a universal symbol system is not formally decidable because the syntactic rules of the language cannot be entirely separated from the physics of the molecules that result from the genetic strings. In evolving symbol-matter systems, the strings constrain the rules just as significantly as the rules constrain the strings. We will see what this means after discussing enzymes.

10.3.3 Molecular Direct Perception

A second well-established fact of life is that the sequential string processing of individual genes ends abruptly at the completion of the primary sequence of a protein molecule. These linguistically terminal strings have the necessary physical constraints to dynamically transform themselves into 3-dimensional structures, by folding, resulting in characteristic biological actions such as self-assembly, allosteric response, pattern-recognition, and selective catalysis. After the terminal string is synthesized, none of the subsequent transformations, recognitions, or actions involves anything that is explained by sequential computation. An enzyme molecule might be usefully compared to a very simple homunculus or to a complex Maxwell demon, but not to a string-processing automaton.

It is not likely that there exists a simpler example of a natural symbol-matter articulation than the gene-enzyme system, but already there are some fundamental questions. First of all, assuming we agree that the gene is a physical string of discrete elements from a small molecular set, how do we justify calling it a symbolic representation? One answer is that it looks and behaves like a part of what we call a language (Pattee 1969). That is certainly one condition, but it is not sufficient for an explanation. We cannot ignore the fact that a physical description of DNA as a double helix with complementary base pairing is a strong contribution to the explanation of its symbolic function. What does this physical description show? It shows that we can ignore an enormous amount of the physical detail and still understand the symbolic behavior. More precisely, our physical description tells us what structures or patterns *the cell* must recognize and what structures are irrelevant for the symbolic reference in the cell. This specific recognition of the symbolically relevant molecular patterns within this mass of other physical detail I would call the cell's primeval perception process.

Another fundamental fact of life is that all primeval perception or pattern recognition and selective action is mediated by enzymes or enzyme-like molecules. This is the case for the cell's sensing of the external environment, for sensing between the cells, and for intracellular recognition of patterns. At this molecular level the recognition-action process is generally pictured as a conditional physical template matching of the enzyme with a target structure, such that if the fit is good enough, a change of shape is induced in the enzyme causing specific, but in a sense gratuitous, physical actions to occur in the target structure, usually involving a specific chemical reaction (e.g., Monod 1971). This mechanical strategy also appears at many higher levels of aggregation (e.g., ribosomes, microtubules, membranes, and so on). I do not see in any of these molecular or aggregate conditional pattern-recognizers and action-executors any dependence on string processing during their actual functioning. At the same time, I do not see any example of a functioning structure that was not syntactically constrained by strings during the initial synthesis of the parts.

10.3.4 The Semantic Closure Principle

Looking more closely at how this comes about in the cell we see that this type of symbol-matter-function dependence is an exceptional kind of interdependence that I call *semantic closure.* We can say that the molecular strings of the genes only become symbolic representations if the physical symbol tokens are, at some stage of string processing, *directly* recognized by translation molecules (tRNA's and synthetases) which thereupon execute specific but arbitrary actions (protein synthesis). The semantic closure arises from the necessity that the translation molecules are themselves referents of the gene strings.

I propose the principle of the semantic closure of symbol and matter as simply a generalization of this primeval fact to higher levels of evolution, including cognitive systems. It can be restated in many ways; (1) the existence of symbolic constraints on physical laws requires selective detection of these symbols; and the selective detection mechanisms require these symbolic constraints or their syntactic productions, in the construction of these mechanisms; (2) in order to say "symbols are processed by syntactic rules" there must be physical embodiments for recognizing the symbol tokens and dynamically executing the rules. The synthesis of these physical embodiments must be constrained through the processing of these symbols.

For example, the semantic closure principle applied to the computer would claim that for computer hardware to be recognized as a symbolic constraint, the physical symbol tokens and rule executors must be constructed through the syntactic productions of the device that ultimately recognizes (i.e., interprets) the symbols. In this case this "device" is of course the human being. The obvious trouble with this semantic closure loop is that it is too long and tangled with other semantic loops to explain. The significant fact is that to use computers we need only make the *artificial* distinction between the syntactic rules and symbols and the hardware as *we* choose. This "we" avoids, or evades, the *natural* tangled semantic closure, and allows the universal symbol system to exhibit a very useful but illusive autonomy.

Thus, the semantic closure principle would logically permit machine self-replication, machine evolution, machine intelligence, and so on, as long as by "machine" we mean a syntactically controlled direct perception device with semantic closure, that is, a machine that works like a cell. On the other hand, neither a pure syntactic machine like the information processors' "physical symbol system," nor a pure dynamical constraint like the direct perceptionists' "smart machine" could evolve or exhibit intelligent behavior, if this principle is valid.

Semantic closure is a physically realistic expression of von Neumann's (1966) logical requirements for a self-reproducing automaton, which he derived before the physical embodiment in cells was known. Semantic closure is, therefore, one condition for evolution by natural selection. In effect, this semantic loop is what defines the "self" in self-replication. Finding a natural process by which a reasonably simple physical system could reach a semantic closure condition is the crux of the origin of life problem, (e.g., Eigen and Schuster 1977), or as it is known in folk cell-psychology, the chicken-egg problem.

10.3.5 Natural Versus Artificial Semantic Closure

One of the great discoveries in the foundations of logic and mathematics was that syntactical precision could be obtained only by relinquishing meaning. However, in the application of mathematics, as in physical theories, the syntactic-semantic distinctions are very difficult to decouple, because in most cases the syntax has evolved only as an abstraction from informal but very meaningful concepts. For example, Euclidean geometry makes a lot of sense no matter how sterile the formal axioms may appear. One could say that natural semantic closure is difficult to break. However, modern digital computers do a very good job of it. It is only through the hierarchy of programming levels that some meaning can be recognized. Most information processors now make this situation clear. A computer is assumed to be in a semantic isolation box with all transducers from symbols to meanings or actions on the outside of the box, and as Newell 1980 explains it, "A prisoner of its input and output domains [p. 148]," therefore, a machine that does not know what it is doing. Computationalists assume that semantic closure could be attained by simply adding the appropriate robotics (i.e., sensors and effectors), at the very beginning and the very end of computation.

The message of cell psychology is that cells don't work that way. Information processing models are as unlike the symbol-matter architecture of the cell as you can get. In cells there is no semantic isolation. Every enzyme recognition and catalysis is a direct semantic act. The cell is a semantically closed network of information processing nodes and direct recognition and action nodes. The information processors' strategy of "once and for all" separation of the semantic and syntactic aspects of cognition does not work for cell psychology. The cell "knows what it is doing" down to the molecular level.

The message of cell psychology does not support direct perception models either, since even though the enzymes do act like "smart machines" with direct pickup of intrinsic, specific information, there is no possibility of explaining such machines without understanding the information processing that led to their synthesis. Ecological realism can not explicitly recognize semantic closure because it does not recognize syntactic processing. However, Gibson's (1979) concept of affordances, and the intrinsic relation of what is "tuned" perceptually to what is "tuned to" in the ecological system, I believe is a recognition of the semantic closure principle using concepts that were intended to make the symbol-matter problem disappear. Given the history of the problem this is a commendable goal. The argument is whether it is a reasonable one.

As Ullman (1980) has pointed out, this argument is over "meaningful decomposition" and "explanatory adequacy." Even at the cell level of simplicity this is not just an empirical problem. We are asking to what extent syntactic rules are meaningfully decomposable into their physical embodiments and to what extent physical detectors are meaningfully decomposable into their syntactic descriptions. It is because of this problem of decomposition that I need to bring some basic physics into the discussion of simple psychologies, not because physics has the answer, but because physics has the same problem. The problem is measurement, and in what sense measurement can be meaningfully decomposed.

10.4 The Physics of Symbols

10.4.1 The Measurement Problem

Measurement is a very restricted form of perception. To measure something means that you are not measuring everything. More formally, a measurement is a mapping from a physical system to a symbol: but the essence of this mapping is the high selectivity or simplification of the system to only the attribute we have chosen to observe. The problem is this: Can we understand the measurement by decomposing the process in detail? To understand *in detail* would put back into the measuring device all the complexity of interaction that the *function* of measurement requires that the device ignores. In other words, a detailed physical description of a measurement process will look just like any physical interaction of two systems. So the physicist needs to know exactly where in this physical interaction (object and measuring device) he can make sure that a symbol appears. The physicist is never sure he has meaningful symbols because on closer analysis he just gets still more physics. Does this sound familiar?

The physicist learned long ago to accept the obvious solution: He does not try to analyze the measurement process, but accepts the result quite literally as a *direct perception* by the measuring device. Most of the "processing time" of the physicist goes into the design and construction of the measuring devices, just as cells spend most of their time in the design and construction of enzymes.

Classical physics does very well by ignoring the process of measurement, and by assuming that whatever is going on in detail it doesn't affect the result. However, the selection of *what* pattern is measured is, of course, crucial. Unfortunately, nature is not that simple. In quantum processes there appears a kind of intrinsic semantic closure between the event and the measurement of the event. This non-classical syntactic dependence in measurement was the motivation for Bohr's Complementarity Principle, which in effect surrenders to the apparent incompatibility of one formal description of both causal laws and measurement rules, both of which are essential for explaining, or even representing, the phenomena. Since in spite of all efforts of the finest minds, there is still no consensus on any alternative view (e.g., Jammer 1974; d'Espagnat 1976), complementarity remains a consistent if not conceptually satisfying mode of thinking. In any case, the complementarity principle does not say anything about *when* physical laws become irrelevant and measurement becomes meaningful.

Some physicists, notably Schrödinger (1961), von Neumann (1955), and Wigner (1965), have passed the buck directly to psychology by arguing that only at the level of consciousness can we be sure that a measurement finally has meaning and is no longer usefully representable by casual laws. My conclusion is that the buck stops much earlier, in fact, at the first *semantically closed,* symbol-matter system like the cell (Pattee 1968). Of course, this conclusion does not resolve the symbol-matter complementarity problem. It only claims that the distinction between causal laws and conventional rules is a problem for life in general, not just for the conscious intelligentsia.

One final intriguing question is whether the evolutionary success of semantic closure itself depends essentially on quantum mechanical measurement processes. I have speculated that this is reasonable, since classically, speed and reliability of rule execution is limited by thermal noise as mechanisms get smaller (e.g., Bennett 1979), whereas a quantum mechanical enzyme may execute measurement in a mode isolated from thermal noise (Pattee 1972).

10.4.2 Laws, Rules, and Representations

The computational view of cognition rests on the possibility of a "principled distinction" between events that can be described "as merely instantiating causal physical or biological laws" and events that require symbolic representations "governed by rules" (e.g., Pylyshyn 1980). If this distinction cannot be made clearly then the computationalist logic falls apart. I believe that classical concepts of physics may be of some value in making this distinction, even though many fundamental questions may still be open. The ecological realist will point out that the conventions of normal physics may not be adequate for explaining perception, which I agree is a possibility. However, until they develop their alternative concepts of physics more clearly I will hang on to the classical rungs of the ladder.

Physical laws are chosen because of very general principles, such as universality and inexorability, as well as very specific principles, such as conservations and invariance to time, position, uniform velocity, and so on (e.g., Wigner 1964). Roughly speaking, to state that something obeys physical law is a tautology, if you understand what the concept of physical law means. Generally our representation of laws of motion is expressed as rate-dependent equations, but to use these equations means applying them to a particular situation. The particular situation is specified by *initial conditions,* obtained by measurements, and other auxiliary conditions called *boundary conditions* or *constraints.* A constraint is the name for a set of complicated boundary conditions.

Constraints are a basic problem because they can always be decomposed in principle to obey only laws and more initial conditions. Constraints are just a *simpler* alternative description of the local situation. For example, a measuring device is one type of constraint. We could, in principle, decompose the device into laws and more initial conditions, but then we would have to make more measurements of all the new initial conditions, thereby losing the function, as we said before.

Rules are not discussed in general by physics because rules can only be executed by *exceptional* boundary conditions or constraints, like measurements, that when described in physical detail become complicated with no improvement in clarity as a rule. Rule-executing constraints are a special case of machine constraint. As Polanyi (1968) has clearly pointed out, a machine constraint has a *biological origin* that cannot in any causal or explanatory sense be attributed to physical laws. Rule-executing constraints may be called natural, like enzymes, or artificial, like transistors, but in both cases they are biologically synthesized. Syntactic rule execution may be interpreted as a special case of measurement where the initial conditions have symbolic content. Syntactic rules are rate-independent, that is, the rate of executing the rule is irrelevant. For example, the rate at which we calculate a function, or the rate at which a protein is synthesized is not syntactically significant (Pattee 1979).

Computational language has many cases of confusing usages of laws and rules. For example, to distinguish between Pylyshyn's (1980) "causal physical and biological laws" and his "rules" requires that we know what he means by biological laws. If he means biological *rules,* then for his argument he needs a principled distinction between the rules that result in direct perception-type measurements and the rules that result in information processing-type measurements. Since cells use both, this distinction is a matter of optimum strategy, not of principle.

Even the concept of "physical symbol system" (Newell 1980) is confusing. Either it is a tautology in the sense that "everything obeys physical laws" or it is begging the question of how symbol systems are explainable as biological rules or physical laws. What Newell actually talks about is more accurately described as artificial, universal, string processing that has nothing to do with physical laws or even machine constraints. On the other hand, I might understand ecological realists writing papers on "physical rule systems" *(symbol* is not in their vocabulary). I would assume they mean that rules are usefully reducible to physical laws according to some "eccentric way of doing science" (Turvey and Carello 1981) and that their functional constraints can be synthesized with no string-processing at any stage (Kugler et al. 1980).

When we come to the ethereal concept of representation, physics has very little to contribute except caution. Cell psychology will probably be more instructive. What does the string of bases in genetic DNA represent? Many geneticists will say it represents the entire individual, or even the species. The gene also represents 3.6 billion years of history. I think these are meaningful and even true statements, but such broad usages do not clarify the concept of representation. The semantic closure principle would suggest that a specific syntactic-semantic loop needs to be established before we can define a representation in any formal sense. For example, we could say that *for the cell* the gene string represents the primary sequence of a protein because we know that the syntactical rules needed for interpreting this string (i.e., synthesizing the protein) are a part of the cell's semantically closed self-replication loop. We could not say that in this loop the gene represents 3.6 billion years of history because the process of interpreting the strings to give this meaning is not specified within this narrow semantic closure. Such a representation can presumably be interpreted within the semantic closure of an evolutionary selection loop, as well as within the closure of an evolutionary biologist's brain, but here we cannot supply the details. The point is that what a given syntactic structure represents depends on the particular semantic closure loop within which it is interpreted. A given syntactic string can, therefore, represent as many properties as there are semantic loops in which it acts as a constraint. The multiplicity of function in evolution as well as the multiplicity of meaning in natural language supports such a multiple-closure concept of representation and meaning.

10.5 Conclusions

Leaving aside our high-level, abstract concept of cognitive activity, and looking at the way the cell works in the simplest concrete terms, we see strings from a small set of elements that are transformed by *rules* into other strings that, in turn, are transformed by *laws* into functional machines with an enormous variety of behaviors. These behaviors include sensing, pattern recognition, amplification, regulation, adaptation, control, and chemical and physical actions of many types. Furthermore, the evidence is very strong that the behavioral mechanisms at all higher levels of aggregation are elaborations of this universal basic process (e.g. Koshland et al. 1982). The contraction of a muscle, the detection by sensory inputs, the recognition of molecular shapes, and even a single synaptic transmission is executed or directly mediated by induced conformation changes in a molecule that was constrained by string processing. These are more or less accepted facts and normal interpretations.

Now back to cell psychology. Are gene strings symbols? Do the molecules have meaning? To answer this question I am proposing a necessary condition on the use of concepts like *symbol, referent,* and *meaning.* Strings have no meaning unless these strings and the dynamics they constrain comprise an empirically traceable process of closure, a syntactic-semantic loop that is self-defining and self-constructing. The cell is the first natural level of semantic closure, and I would say that the cell's genes have a primeval symbolic function and meaning for the cell. The same genes may evolve additional meanings for the multicellular organism, and for any higher levels where semantic closure exists. At higher evolutionary levels the products of genes can become symbol tokens themselves within semantically closed epigenetic loops. At the level of the brain, this principle of the closure of direct perception and information processing is the key idea of Neisser's (1976) *perceptual cycle* in which he compares the active schema that prepares perceptions to the genotype. What multicellular psychology needs is simpler cases of epigenetic semantic closure where the symbolic molecules and dynamic aggregates can be empirically distinguished.

Semantic closure gives a new perspective to the idea of perceptual and symbolic *primitives.* To a large degree, the choice of primitives is what distinguishes a conceptually powerful explanation from a mere description or "imitation game." Thus, the information processors must constrain their universal symbol systems that can imitate anything, by assuming the existence of natural primitives which they view as a kind of fixed biological hardware. This "functional architecture" is emulated by programming a functionally similar "virtual machine" architecture. Similarly, the direct perceptionists need to constrain the universality of ecological physics that everything obeys, by assuming the existence of exceptional constraints which they also view as natural primitive hardware, but hardware of such exquisite design that no information processing is necessary. Cell psychology and semantic closure are not consistent with decompositions into these primitives. Any symbolic "primitive" in real life is itself decomposable, since its expression is controlled by a set of direct recognition-action molecular machines. Similarly, any direct perception "primitive" in real life is itself decomposable, since its construction is controlled by a set of symbolic molecular strings.

The relevance of cell psychology to higher levels of behavior remains largely an empirical question. Evolution is opportunistic. I would surely be surprised if the brain did not use some string processing to solve string processing problems, and some direct pattern recognition to recognize patterns. However, cell psychology shows us a successful evolutionary strategy of processing symbol strings to construct and control direct recognition-action machines at higher and higher levels of organization. I would also be surprised if the learning of skilled pattern recognition and skilled action at the highest levels did not take advantage of this very general strategy.

References

Bennett, C. H. (1979). Dissipation-error tradeoff in proofreading. *Biosystems, 11*, 85–91.

Carello, C., Turvey, M. T., Kugler, P. N., & Shaw, R. E. (1982). Inadequacies in the computer metaphor. In M. S. Gazzaniga (Ed.), *Handbook of cognitive neuroscience*. New York: Plenum.

Davidson, D. (1981). The material mind. In J. Haugeland (Ed.), *Mind design*. Cambridge, MA: MIT Press.

d'Espagnat, B. (1976). *Conceptual foundations of quantum mechanics* (2nd ed.). Reading: W. A. Benjamin.

Eigen, M., & Schuster, P. (1977). The hypercycle. *Naturwissenschaften, 4*, 541–565.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Hoffman, G. (1975). The stochastic theory of the origin of the genetic code. In H. Eyring, C. J. Christianson, & H. S. Johnson (Eds.), *Annual reviews of physical chemistry* (Vol. 26). New York: Academic.

Jammer, M. (1974). *The philosophy of quantum mechanics*. New York: Wiley.

Koshland, D. E., Jr., Goldbeter, A., & Stock, J. B. (1982). Amplification and adaptation in regulatory and sensory systems. *Science, 217*, 220–225.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. Amsterdam: North Holland.

Minsky, M. (1980). Decentralized minds. Commentary on J. R. Searle, minds, brains, and programs. *Behavior and Brain Sciences, 3*, 417–457.

Monod, J. (1971). *Chance and necessity*. New York: Knopf.

Neisser, U. (1976). *Cognition and reality*. San Francisco: W. H. Freeman.

Newell, A. (1973). You can't play 20 questions with nature and win. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic.

Newell, A. (1980). Physical symbol systems. *Cognitive Science, 4*, 135–183.

Pattee, H. H. (1968). The physical basis of coding and reliability in biological evolution. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Prolegomena, Vol. I). Edinburgh: Edinburgh University Press.

Pattee, H. H. (1969). How does a molecule become a message? *Developmental Biology, Supplement, 3*, 1–16.

Pattee, H. H. (1972). Physical problems of decision-making constraints. *International Journal of Neuroscience, 3*, 99–106.

Pattee, H. H. (1979). Complementarity vs. reductionism as explanation of biological complexity. *American Journal of Physiology, 5*(3), R241–R246.

Polanyi, M. (1968). Life's irreducible structure. *Science, 160*, 1308–1311.

Pylyshyn, Z. (1980). Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences, 3*, 111–169.

Schrödinger, E. (1961). *Science and humanism*. Cambridge: Cambridge University Press.

Turvey, M. T., & Carello, C. (1981). Cognition: The view from ecological realism. *Cognition, 10*, 313–321.

Ullman, S. (1980). Against direct perception. *Behavioral and Brain Sciences, 3*, 373–415.

von Neumann, J. (1955). *Mathematical foundations of quantum mechanics*. Princeton: Princeton University Press.

von Neumann, J. (1966). *Theory of self-reproducing automata.* Edited and completed by A. W. Burks. Urbana: University of Illinois Press.

Waters, D. (1982). Syntax of the genetic code. *Journal of Theoretical Biology* (submitted)

Wigner, E. P. (1964). Events, laws, and invariance principles. *Science, 145*, 995–999.

Wigner, E. P. (1965). Remarks on the mind-body problem. In I. J. Good (Ed.), *The scientist speculates*. New York: Capricorn Books.

11 Universal Principles of Measurement and Language Functions in Evolving Systems

The ability to construct measuring devices and to predict the results of measurements using models expressed in formal mathematical language is now generally accepted as the minimum requirement for any form of scientific theory. The modern cultural development of these skills is usually credited to the Newtonian epoch, although traces go back at least 2,000 years to the Milesian philosophers. In any case, from the enormously broader evolutionary perspective, covering well over three billion years, the inventions of measurement and language are commonly regarded as only the most recent and elaborate form of intelligent activity of the most recent and elaborate species.

In this discussion I argue that such a narrow interpretation of measurement and language does not do justice to their primitive epistemological character, and that only by viewing them in an evolutionary context can we appreciate how primitive and universal are the functional principles from which our highly specialized forms of measurement and formal languages arose. I present the view that the generalized functions of language and measurement form a semantically closed loop which is a necessary condition for evolution, and I point out the irreducible complementarity of construction and function for both measuring devices and linguistic strings. Finally, I discuss why current theories of measurement, perception, and language understanding do not satisfy the semantic closure requirement for evolution, and I suggest approaches to designing adaptive systems which may exhibit more evolutionary and learning potential than do existing artificial intelligence models.

My approach is to generalize measurement and linguistic functions by examining both the most highly evolved cognitive systems and the simplest living systems that are known to have the potential to evolve, and abstracting their essential and common measurement and linguistic properties. I want to emphasize that when I speak of molecular language strings and molecular measuring devices I am not

---

Reprinted from *Complexity Language and Life: Mathematical Approaches*, John Casti, and Anders Karlqvist, Eds. Berlin: Springer-Verlag, 1985, pp. 268–281.

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7,
DOI 10.1007/978-94-007-5161-3_12,

constructing a metaphor. Quite the contrary, I mean to show that our most highly evolved languages and measuring devices are only very specialized and largely arbitrary realizations of much simpler and more universal functional principles by which we should define languages and measurements.

11.1 Generalized Measurement

The classical scientific concept of measurement requires a distinct physical measuring device that selectively interacts with the system being measured resulting in output that has a symbolic interpretation, usually numbers. Most scientists regard the output of a number as an essential requirement and indeed numbers are required if the language of science is restricted to mathematics. If the laws are expressed by equations of motion, then the initial conditions must be numbers if we are to use the equations to predict other numbers. However, without questioning the enormous advantages of numbers and formal mathematical representations of laws, it is obvious that measurements are possible without numerical outputs (e.g., Nagel 1932). For example, timing, navigating, surveying, weighing, and even counting were once accomplished by iconic, mimetic, or analog representations. Today the trend is away from the outputs of traditional laboratory measuring devices with visible numerical scales and toward transducers that feed computers and robots directly. In all cases the type of output from a measurement is chosen according to the particular functional requirements of the system as a whole.

The essential point is that while the selection of *input patterns*, the choice of *output actions*, and the *relation* of input to output in any measuring device is largely arbitrary, the only fundamental requirements for useful measurements are the *precision and reproducibility*, or *local invariances, of the input-output relation*, and the *functional value* of the entire operation to the system doing the measuring. The requirement of reproducibility means that the measuring device must be *isolatable* from the system being measured, and *resettable*, so that the measurement process can be repeated an arbitrary number of times to give the same output for the same input pattern. However, such an abstract description of measurement is incomplete, since it omits the crucial requirement of *system function*, or the *value* of the measurement.

From the abstract definition of measurement alone we would conclude that any relatively fixed or constrained set of particles in a physical system qualifies as a measuring device if we interpret pattern as simply the initial conditions of the free particles and action as the alteration of their free trajectories after collision with the constrained set. Thus, we might say that a rock in a stream maps the input flow pattern to the output action of turbulence, or say that in crystal growth the constraint of a dislocation on a crystal surface maps the patterns of molecular collisions to the specific action of binding more of its own constituents. However, we do not normally call these cases measurement processes. By contrast, the pattern recognition required for specific substrate binding and catalytic action of cellular enzymes I would call a measurement, even by the most rigorous definitions that apply to highly specialized, artificial devices. How do I justify this? Clearly, the enzyme's action is more complicated than crystal growth, but I do not see the level of complexity of the measuring device as the only criterion; for example, calipers are a simple, artificial constraint that we may use to measure size. The only distinction I find convincing is that of *system function* or, more specifically, that of pattern-action mapping that supports the persistence or survival of the system, and subsidiarily of the measuring constraints that make up the system. In other words, there must be *functional closure.* It is necessary that the enzyme serves a function in the cell for its pattern recognition and catalytic action to be called a measurement. This is still too broad a definition, since it gives no clues as to the characteristics of function, other than survival, that are required of measurement. We must specify some further conditions on this mapping from patterns to actions that are necessary for efficient or effective measurements. Are there also conditions on the way that successful *systems* of measuring devices interact? Let us consider what is common to some extreme examples of successful measuring devices.

11.2 Measurement as a Classification

The most important, and yet the most deceptive, aspect of our highly evolved artificial measurements is the feeling we have as intelligent observers that we know what attribute we are measuring independently of the measuring constraints. This is a half-truth. We usually have an abstract concept of what attributes we wish to measure, and design the constraints of the measuring device so that its output action expresses these attributes and minimizes all others. Since the output action is designed to be very simple, we often tacitly assume that the corresponding input patterns are very simple. For example, we think of temperature as a simple property of a gas, but our thinking does not change the complex molecular collisions of the gas. This is actually a useful deception in building classical models, although it leads to erroneous results in quantum mechanics. In fact, the measuring device necessarily interacts physically with all of the system's innumerable degrees of freedom, and it is precisely because of the innumerable internal constraints of that particular device that only a few degrees of freedom are available for the output actions.

It is primarily this property of mapping *complex input* patterns to *simple output* actions that distinguishes useful measurement functions from merely complex physical constraints. Without this complex-to-simple or many-to-one mapping process we would not be able to identify equivalence classes of events and consequently we would not be able to construct simple models of the world. I would go further and claim that *this classification property of measurement is an epistemological necessity.* Without classification, knowledge of events would not be distinguished from the events themselves, since they would be isomorphic images of each other. This also implies symmetry in time, and measurement must be an irreversible process.

From a broad biological perspective, the entire nervous system has evolved for the principle function of quickly and reliably mapping the ineffably complex configurations and motions of the environment to a very few vital actions; that is, run, fight, eat, sleep, mate, play, etc. Although these actions can be decomposed into complex subroutines, the decision is still which of only a small number of actions to employ. The entire organism can therefore qualify as an extreme case of a generalized measuring device. Let me return now to the other extreme of evolution and consider measuring devices at the molecular level.

At the cellular level we have the example of the single enzyme molecule. The action of an enzyme, like the action of an artificial measuring device, may be described very simply. Generally, it is the catalysis of one particular covalent bond and, consequently, we might think of the corresponding input pattern simply as one particular substrate molecule. But this would miss the essential property of an effective measuring device, which is to reduce the complexity of input interactions by means of its internal constraints. When we speak of an enzyme as highly specific it is another way of saying that it recognizes or distinguishes very complex input patterns.

This ability to recognize complex input patterns and, as a consequence, execute a simple action requires physical constraints of a special type. Since the many-to-one mapping is arbitrary, the constraints must arbitrarily couple the *configurations* available for fitting the input pattern to the *motions* of the device that produces the output actions. In physics these are called nonholonomic or nonintegrable constraints. A holonomic constraint is a restriction on the configuration of a set of particles, such as occurs in forming a crystal from a solution of molecules. This freezing-out of configurational degrees of freedom necessarily freezes-out the corresponding motions of the crystallized molecules, so that we see the constrained system as a rigid solid. A nonholonomic constraint may be defined as a restriction on the motions of the particles *without* a corresponding restriction in the particle configurations. In other words, a formal expression of a nonholonomic constraint appears as a peculiar equation of motion for selected velocity components, where certain configurational variables serve as initial conditions. However, we cannot generally eliminate any configurational variables of the system by using these relations because of the nonintegrability of the equations of constraint. This results in a flexible or allosteric configuration. What we call machines are made up of holonomic, rigid parts that are coupled by nonholonomic, moving linkages. In such machines more configurations of the parts are allowed kinematically than are allowed in the dynamic motions of the parts (e.g., Pattee 1972b). In proteins it is these nonholonomic constraints that couple the complex configurations or patterns of the substrate to the allosteric motions causing catalytic actions.

The complexity of patterns that can be usefully distinguished clearly depends, in part, on the complexity of the internal constraints of the measuring device that fits the pattern. What is not so clear, but equally important for recognition, is that the output action must be simple and repeatable. In fact, we can imagine a complex fit that requires complex constraints without any corresponding simple action, as in dumping a pile of gravel. We also speak of complex actions resulting from complex constraints, as in the weather. But it is only when complex interactions result in simple, repeatable actions that we speak of recognizing patterns. Enzymes require hundreds of amino acid residues to fold into a structure which we say fits the substrate; but *any* solid has molecules that physically fit each other just as well, yet we do not generally picture solids as pattern recognizers. It is only the simple catalytic action that establishes the fit interaction as a pattern candidate; but I would again argue that the only objectifiable existence of patterns is ultimately established by some form of system closure. That is, *the distinguishing property of measurement constraints* is *that their pattern-action mapping supports the system that is necessary to synthesize these constraints.* Since this is such a fundamental condition, let me discuss it in more detail. We shall see that for evolution to be possible, functional closure must be more complex than just autocatalytic cycles.

11.3 Function Requires Construction

Returning now to the human level we can say that the primary function of measurement is to map the ineffably complex interactions of the physical world into attributes which are necessary for our survival in this world. To realize this function, it is obviously necessary for us to pay attention to these attributes. This justifies the epistemological illusion of thinking about the world in terms of these measured attributes; that is, in terms of the simple *outputs* of the measuring devices rather than the complex inputs. In the everyday use of observations and measurements there is no survival value in analyzing the inner details of measuring devices. In other words *performance* of measurements does not benefit from *analysis* of the constraints of the measuring device.

In fact, if one analyzes the measurement constraints using a microphysical description, the measurement *function* unavoidably disappears into a measurement-free physical system with more degrees of freedom. On the other hand, *it* is *from this more detailed physical system that the complex measurement constraints must have been synthesized in the first place.* This means that we must have *control* over physical details of constructing measurement devices even though we do not want or need knowledge of these details while we actually perform measurements. The measurement activity therefore requires both *functional primitives,* in the sense that any analysis of the constraints of the measurement device necessarily obliterates the essential classification action, and *constructional primitives,* in the sense that knowledge of the function of the device can result in no necessary rules for synthesizing the device's constraint.

This apparently improbable interrelation between genes and enzymes is the simplest case of what I call *semantic closure* (Pattee 1982). By general semantic closure I mean the relation between two primitive constraints, the generalized measurement-type constraints that map complex patterns to simple actions and the generalized linguistic-type constraints that control the sequential construction of the measurement constraints. The relation is semantically closed by the necessity for the linguistic instructions to be read by a set of measuring devices to produce the specified actions or meaning. The semantic closure principle is supportable from several levels:

1. As an empirically based generalization from the facts of molecular biology.
2. As a theoretical requirement based on the logic of heritable systems (e.g., von Neumann 1966; Polanyi 1968).
3. As an epistemological condition necessary for the distinction between matter and symbol (Pattee 1982).

It may also be stated, as a complementarity principle, that the properties of measurement and language cannot be adequately defined individually, but form an irreducible, complementary pair of concepts.

11.4 Generalized Language

There is common agreement on many of the universals of language *structure* (e.g., Hockett 1966). Natural and formal languages are discrete, one-dimensional (1-D) strings of elements from a small alphabet. The strings are further constrained by lexical and syntactic rules which may be very simple or very complex. These rules may be precise and explicit, as in formal languages, or ambiguous and difficult to formulate as in natural languages. Language strings are constructed and read sequentially, although all natural languages also have the essential metalinguistic ability to reference themselves out of sequence; that is, to construct strings that refer to other strings in the language. From what is known of the structure of the gene it appears to qualify fully as a natural language system (e.g., Pattee 1972a).

When it comes to language *function* it is more difficult to find simple generalizations, let alone common agreement. Language unquestionably has many functions; for example, memory, instruction, communication, modeling, thought, problem-solving, prediction, planning, etc. What I am proposing is not inconsistent with any of these functions. However, my criteria for functions in both measurement and language are based on the most *primitive* conditions for evolvable systems. These include:

1. The ability to construct and coordinate measuring devices and other functional structures under the control of a heritable description (i.e., genetic control).
2. The ability to modify function by changing the description (i.e., mutability).
3. A heritable process for evaluating the description-construction system as a whole (i.e., natural selection).

The impressive techniques of molecular biology have shown us in some detail how present cells accomplish these processes, so in a phenomenological sense they are no longer considered problems by biologists. However, there remain the essential mysteries of how such cellular systems came to exist and how multicellular systems develop. That is, how does such a coordinated set of linguistic instructions and measuring constraints evolve from a nonliving physical world and how are such intricate multicellular morphologies constructed and maintained by these linguistic and measurement devices? I comment on approaches to these problems in the last section but, since I have no solutions, for the present I assume the existence of cells and simply generalize from the structure of linguistic constraints and how they function in the cell system.

We considered the basic function of measurement devices as mapping complex patterns to simple actions. In cells, these actions are typically the catalysis of a single bond; in effect, the smallest change that can be made in the constraints of a molecular system. However, this small change is only made if a complicated set of other constraints is satisfied, namely the recognition of the substrate molecule. In a linguistic device the functions are quite different. The function of the linear sequence is to control the sequence of actions necessary to construct the measuring device, but it does this through the simplest possible type of constraint, the chain of single bonds. At the other extreme of language constraints, we find that one principle function of our spoken and written languages is to give instructions; and it is an impressive fact, often taken for granted, that by forming discrete strings from about 30 types of simple marks we can effect the construction of almost any conceivable pattern, whether it is in the brain, in the actions of the body, or in the construction of artifacts. How these transformations take place from the simple, 1-dimensional string of constraints to the physical structures and actions represented by these strings is almost a total mystery for natural language, even though we find we are able to know the meaning from the strings. But by contrast, at the molecular level we know in great detail how the genetic strings are transformed into the structures and actions represented by the strings, but we have no way of deriving the meaning of any string; that is, of how to tell from the genetic message alone what the function is of the protein it describes.

A generalized language might therefore be characterized as a simple chain of constraints that controls the construction of complex patterns. If we try to formalize this further, as we did with measurement, we might be tempted to say that linguistic devices map a domain of 1-dimensional constraints to a range of n-dimensional patterns. However, this would be a misleading abstraction. In the case of the measuring device, it is the actual constraints of the device itself that recognize the input pattern and are physically responsible for the output action. Therefore, by saying that the device *maps* the input pattern to output action we mean that it is responsible for dynamically *executing* the mapping. On the other hand, a string of constraints in a language is dynamically inactive. Language strings are pure configurations; that is, they have no significant motions or velocity components. Thus, symbol strings are rate independent in the sense that their meaning, or what they control, does not depend on how fast they are read.

11.5 Semantic Closure

We explained earlier how the *action* of measurement constraints is functionally primitive, since analysis of the details of the constraints interferes with the measurement function. In a similar way the *meaning* of a linguistic string is functionally primitive, since analysis of the mechanisms of production of the string interferes with the meaning. In practice, we look at the results of a measurement and do not confuse ourselves with the constructional details of the measuring device. Similarly, when we generate linguistic strings we focus on the meaning, not the mechanics of production.

This complementary primitiveness of measurement and linguistic meaning is not only an observable fact of biology but also, I believe, an epistemological necessity. As we said earlier, it is essential that we be able to directly picture the world from what we perceive or from the outputs of measuring devices without having to also know the physical details of the perceiving or measuring constraints as parts of the non-measuring interactions of the physical world. This requirement of semantic primitiveness of perception and measurement accounts for what I call the epistemic illusion of the reality of the world, which is not involved with the complex and largely arbitrary constraints that execute perception and measurement. The alternative possibility, that we must analyze these measurement constraints, only leads to irrelevant details at best or an infinite regress at worst. In a complementary sense it is essential that we be able to directly grasp the meaning of linguistic strings without becoming involved with the complex and largely arbitrary details of the constraints that generate and interpret strings. Just as in the case of measurement, this requirement of semantic primitiveness of language accounts for the epistemic illusion that strings have an intrinsic meaning independent of the dynamical constraints that generate them or that they ultimately control. Only through semantic closure do these two primitives complement each other and form an autonomous, evolvable system. The semantic closure principle allows us to treat the *action* of a measuring device as primitive because the details of its construction are accounted for by a linguistic string, while the *meaning* of the linguistic string can be treated as primitive because the details of interpretation are accounted for by a set of measuring devices. For me, it is this fundamental relation between the *relative primitives* of measurement and language constraints that distinguishes evolvable or epistemic systems from normal physical systems. In the final sections I elaborate on why this closure principle offers more promise for models of evolvable systems than other approaches. Before doing this, let me summarize the properties of generalized measurement and language.

11.6 Properties of Generalized Measurement

1. Measuring devices are localized, isolatable, resettable structures with repeatable actions.
2. Measuring devices have no intrinsic output actions, but may be triggered to simple actions by specific input patterns (nonholonomic constraints).
3. Measurement constraints obey all physical laws, but are not derivable from laws (generated by system function).
4. Measuring devices are constructed sequentially under the control of linguistic constraints, but a complete, finite set of measuring devices is necessary to read linguistic strings (semantic closure).

5. Measuring devices execute a many-to-one mapping from complex input patterns to simple output actions (classification).
6. Measuring devices do not occur in isolation, but form functional, coherent sets within a system.
7. The value and quality of any measurement is a system property determined by the survival of the system in which it functions.
8. Beyond these properties, the domain of input patterns, the range of output actions, the choice of mapping, and many other aspects of measuring devices are largely arbitrary.

11.7 Properties of Generalized Language

1. Language structures are discrete, 1-dimensional strings made up of a small number of types of elements.
2. Language strings have no intrinsic actions, but may trigger action in measuring devices (nonholonomic constraints).
3. Linguistic strings obey all physical laws, but are not derivable from laws (generated by system function).
4. A complete, but finite, set of measuring devices is necessary to read and interpret linguistic strings.
5. Linguistic instructions are necessary to control the synthesis of this interpreting set (semantic closure), as well as the synthesis of other functional components of a system.
6. Language strings are transcribed sequentially, independently of rate, but they may reference themselves out of sequence (metalanguage).
7. The value and meaning of any linguistic string is a system property determined by the survival of the system that it controls.
8. Beyond these properties, the physical structure, the choice of alphabet, the units of meaning, and many other aspects of language strings are largely arbitrary.

11.8 Models of Evolution

How can these primitive closure requirements for measurement and language be incorporated into a model of an evolving or learning system? How would such a model differ from previous models? Let us begin with the second question. Many more or less literal simulations of genetically controlled, self-reproducing systems have been studied, beginning with von Neumann's self-reproducing automaton in which he first explicitly recognized the need for a genetic description as well as a universal constructor that must read and execute this description if evolution is to produce increasingly complex systems. However, von Neumann (1966) was more interested in the logical or linguistic aspects of the model than in the physical aspects of pattern recognition and measurement. He was well aware of this neglect of the physical aspects of the problem ("one has thrown half the problem out the window and it may be the more important half"), but at that time (ca. 1948) the Turing concept of computation was well-developed, while molecular biology was still a great mystery.

Many later simulations of evolution have been attempted for the purpose of improving the adaptation or optimization process in formal or artificial systems (e.g., Bremermann 1962; Fogel et al. 1966; Klopf and Gose 1969; Holland 1975; Barto 1984). Only a few models of evolution have been constructed to help conceptualize and test the postulates of neo -Darwinian theory (e.g., Moorehead and Kaplan 1967; Conrad and Pattee 1970). In all but one of these models, the process of natural selection is accomplished by fitness criteria which are explicit and pre-established by the programmer. In the Conrad and Pattee model no explicit fitness criteria were introduced. Instead, a set of general conditions or rules of interaction between organism and environment were defined, such as conservation of metabolic resources. However, the nature of the environment with respect to the organisms was pre-established; that is, no genetically modifiable measurement constraints were introduced in this model. Thus, in the existing models of evolution the environment has been represented as a fixed, objective framework that produces the selection pressures on the populations of organisms. Our present complementary view of language and measurement requires the epistemic condition that the organism can only respond directly to the simple output of measurements of the environment. As we have seen, these simple outputs are a consequence of complex constraints resulting from genetically controlled syntheses. However, there is no explicit relation of the gene string to the input-output mapping of the measuring device. Gene strings that construct measuring devices cannot be thought of as programs that manipulate data structures in a computer. In the latter case, every program instruction must be completely explicit. Explicit actions require that all types of inputs, outputs, and hardware operations be pre-established. By contrast, in the organism it is the genetic instructions that construct the hardware that determines all the inputs, outputs, and actions. Genetic consequences are therefore entirely implicit. One cannot assign an element of the gene to an element of action, yet this is the central requirement of a program or effective procedure in computation. Furthermore, simply to say that the architecture of present computers is totally unlike the architecture of organisms is a misleading understatement, since even the concept of architecture plays an entirely different role in organisms to that in computers. For these reasons any form of computational metaphor for organisms must be treated with skepticism.

Up until quite recently the predominant view of genetic control has been very much like the view of computation as an explicit program control of data strings in memory. The alternative view that morphogenesis depends both on autonomous dynamics (archetypes) and internal constraints (chreods) for which genes provide only local switching forces is well known (Waddington 1968), but for many years lacked empirical evidence and a conceptually clear, formal model. Currently, such topological and dynamical models of morphogenesis are more popular largely because of the application of elegant mathematical formalizations of the singularities, bifurcations, degeneracies, and instabilities of dynamical systems. These mathematical and physical theories of continuous systems arose from completely distinct concepts and methodologies to those of the computational models of morphogenesis, yet they have also led to models for the growth of many types of biological patterns as well as impressive claims for more general powers (e.g., Thom 1975; Prigogine 1980; Eigen and Schuster 1979; Haken 1981). However, in spite of these significant contributions to mathematical and physical theory, biologists usually perceive the excitement over these formal models as coming more from the physicists and mathematicians who are impressed with the complex patterns that can be generated from such simple equations and boundary conditions. The problem is that molecular genetics is itself so well-established at the foundations of biology that dynamical models are not likely to be useful until they can incorporate these linguistic constraints into their models of evolution and development. While some of these dynamical models have helped clarify measurement constraints (e.g., Prigogine 1980), none of them has directly contributed to the genotype-phenotype closure relation that is necessary for evolution.

It is also instructive to review current theories of cognitive activities at the other end of the evolutionary scale where the subjects of interest are perception, action, learning, language, knowledge, and other forms of intelligent activity. It is significant that here also we find two opposing schools, one based on explicit linguistic strings and the other on implicit measurement dynamics. The first school arose from logic and computation theory, and is now dominated by the paradigm of the computer as the universal symbol system that can model cognitive tasks such as pattern recognition, classification, learning, and understanding natural language. It is the claim of the computationalists or information processors that these tasks can be understood as purely linguistic or string-processing activities without reference to measurement or any physical dynamics, except as pre-established input and output transducers for the strings. These computational modelers appear to have a principled commitment to the epistemic illusion characteristic of linguistic constraints that strings contain implicate meaningful information, and that by processing these strings with a sufficiently clever rewriting of the rules, this meaning can be explicated (e.g., Newell 1980; Pylyshyn 1980). In a somewhat less principled way, the information processors are committed to the complementary epistemic illusion of measurement that only the simple output action need be entered into their models, and that the origin of the complex dynamical constraints that generate these simple outputs need not be considered as a part of their cognitive process.

The opposing school, which arose from the ecological physics approach of J. J. Gibson (1979), takes the other extreme of basing their models on a principled avoidance of linguistic constraints, which they argue are neither essential for mapping perception to action nor for the construction of measurement constraints. Ecological physics models are based on extensions of the dynamical singularity theories of physics (e.g., Turvey and Carello 1981; Turvey and Kugler 1984), and understandably emphasize perception-action models rather than genetic control or language understanding.

Both the information processing and the ecological physics schools of cognitive modeling appear to have committed themselves to their exclusive methodological principles without serious consideration of the empirical facts of development and evolution. In effect, the information processors are committed to the principle that discrete strings possess intrinsic meanings independent of the physical dynamics that generate the strings, while the ecological physicists are committed to the principle that physical dynamics possess intrinsic meanings independent of the genetic strings that have constructed the dynamical constraints. One simple, but very frustrating, fact of evolution is that natural selection does not follow the physical or logical principles of most other scientific models, but operates only through opportunistic and even haphazard experiments. Survival depends on balancing many highly interrelated qualitative system properties such as speed, reliability, efficacy, recovery from error, efficiency, and adaptability. Thus, although it may be technically efficient for us to recognize shape by computation on a string of data obtained by an arbitrary scanning of the shape, the enzyme is much quicker using direct 3-dimensional template recognition with no computation whatsoever; and although it is technically possible to cast a machine from a 3-dimensional template with no string processing, the enzyme is constructed more reliably by sequentially processing a gene string. At the cognitive level why should this opportunistic strategy be different? We can recognize the number of rocks in a pile directly if there are less than 6 or 7, but we must count them sequentially if there are more. In a fraction of a second we directly recognize our complex friends in a crowd, but may have to follow long strings of inductions to identify a simple mineral in a rock. The brain, like the cell, has clearly evolved the power both to directly perceive patterns (measure) and to process strings (compute).

To me, the effort to model the brain as exclusively one or the other type of constraint may be useful engineering—in principle it can be done—but that is not our problem. Our problem with the nervous system is to understand the functional interrelation of direct perceptions and language necessary for efficacious action and learning, just as the problem with the cell is to understand the functional interrelations of gene strings and cellular dynamical constraints necessary for development and evolution. These interrelations are certainly very complex and largely unknown, but what is perhaps the most fundamental evolutionary fact we already know, and that is the meaninglessness of strings or dynamics taken in isolation. From the evolutionary perspective it is only the semantic closure of genotypic language strings and phenotypic measurement dynamics that defines any biological organism in the first place. Whether any physical strings or dynamical constraints can be said to form a language or a measuring device, or whether either has function or meaning can only be decided in terms of its origin and function in the life of the organism.

11.9 Conditions for Artificial Evolving Systems

I now come to the question of how this semantic closure property of measurement and language can be incorporated into an artificial system. Although language and measurement are complementary primitives they do not relate symmetrically. We pointed out that measurement constraints are dynamically active without linguistic inputs, even though they may have been constructed under linguistic constraints.

Measurement devices physically execute the mapping from input patterns to output actions. This means that a system of measuring constraints, once constructed, can perform complex dynamical tasks without further linguistic control. In other words, the specific actions of measurement systems do not require a program to run them. By contrast, a linguistic constraint has no intrinsic dynamics; it is rate independent, and it therefore can execute no rule or action by itself. Every action of a linguistic system must therefore have an external rule or program step to execute it. In effect this is how computation is defined. Only a string that is mapped into another string *by means of* an effective procedure can qualify as formal computation; but a measurement by itself is not an effective procedure since it has no explicit input. Conrad and Hastings (1985) have proposed naming such direct transformation a new computational primitive, but since there is no explicit input, they must use a nonstandard definition of computation. Gibsonians often refer to measurement constraints as "smart machines" that accomplish their function without computation, to contrast them with string processing that requires smart programming if any useful output is to result.

I am proposing that any model of an evolutionary process must clearly represent and functionally distinguish language and measurement constraints (i.e., the genotype and phenotype) and must preserve the properties and relations of each. This includes the construction of the measuring devices under the constraints of the linguistic strings and the reading of these strings by measuring devices. It must also include the ability of the strings to gradually or suddenly modify the inputs, outputs, and mappings of the measuring devices and must allow the representation of measuring devices to function under an autonomous dynamics once they have been constructed. This latter condition is difficult to fulfill in an artificial model since the function of a measuring device depends on its interactions with an environment. If we try to simulate the natural environment, the model becomes very complex and yet is incomplete. On the other hand, if we invent too simple an artificial environment, the measurement mapping becomes trivial. The engineering approach is to have the model adapt to the real natural environment, but this requires the construction of real measuring devices under genetic control. This may be practical, but one could question its status as an explanatory model, or even a model at all, since it would appear to be a real evolving system. One more pedagogic-type model might utilize an artificial environment that could be gradually modified in the hope of inducing new measurements by the organism. What are the simplest conditions under which we can expect such emergent behavior?

It appears obvious that the simulation of language constraints on a computer is simpler than the simulation of measurements. However, there is an enormous difference between natural languages and artificial programming languages, which is easily recognized, but not understood. Typically, computer languages do not tolerate mutations or recombinations, whereas genes and natural languages depend on such changes for evolution and creative expression. One difference which may be significant is the lack of complementary measurement constraints in current computer architectures. Since linguistic constraints have no intrinsic dynamics, the computer does nothing unless given a program step. Furthermore, this step must be explicit;

that is, the mapping from the domain of program steps to the range of output actions must be unconditionally defined in advance. This total dependency on linguistic inputs results in a total intolerance to the absence of inputs or to inputs with syntactical error. It also follows from the requirement of explicitness in the program steps that errors are also explicit; that is, changes in input-output mappings cannot be gradual. Natural systems, on the other hand, operate with measurement constraints under autonomous dynamics that do not require linguistic inputs for their function. Furthermore, this function depends only implicitly on the linguistic strings that controlled their construction; that is, the mapping from strings to measurement function cannot be specified as a sequence of unconditionally defined steps as in a program. Each linguistic step contributes to the final function only in conjunction with the contributions of other steps so that no single linguistic input step can be assigned an unconditional consequence in the output action. This input-output relation can be observed most directly in the folding transformation that converts the linguistic string constraints of the polypeptide's primary structure into the 3-dimensional globular structure of a functioning enzyme. The significant result of this transformation is that a mutation or recombination of the linguistic string may result in all degrees of functional change, from virtually no change, to gradual or continuous change, to discontinuous change, to a new function. This same variability in meaning occurs in natural language where a single change in a letter or word may result in no change of meaning, a shift of meaning, or an entirely new meaning.

The nature of this relation between description and function or between language and meaning is certainly the most crucial and yet the most puzzling aspect of any epistemic or evolutionary system. It is a problem as old as philosophy, and even now it is not clear that a complete explanatory model is possible. My only conclusion from this discussion is that unless an artificial system contains representations of the constraints of both generalized language and generalized measurement as well as the complementary relations between them that I have described as *semantic closure*, the model is not likely to evolve like living systems or to contribute significantly to the theory of evolution.

References

Barto, A. (1984). *Simulation experiments with goal-seeking adaptive elements. Final report, June 1900–August* 1988. Ohio: Avionics Lab., Air Force Wright Aeronautical Lab., Wright-Patterson Air Force Base, 45(33).

Bremermann, H. J. (1962). Optimization through evolution and recombination. In M. C. Yovits, G. T. Jacobi, & G. D. Goldstein (Eds.), *Self-organizing systems*. Washington, DC: Spartan Books.

Conrad, M., & Hastings, H. M. (1985). Scale change and the emergence of information processing primitives. *Journal of Theoretical Biology, 112*(4), 741–755.

Conrad, M., & Pattee, H. H. (1970). Evolution experiments with an artificial ecosystem. *Journal of Theoretical Biology, 28*, 393–409.

Eigen, M., & Schuster, P. (1979). *The hypercycle: A principle of natural self-organization*. Heidelberg/Berlin/New York): Springer.

Fogel, L. J., Owens, A. J., & Walsh, M. J. (1966). *Artificial intelligence through simulated evolution*. New York: Wiley.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Haken, H. (1981). Synergetics: Is self-organization governed by universal principles? In E. Jantsch (Ed.) *The evolutionary vision*, AAAS Selected Symposium 61. Boulder: Westview Press.

Hockett, E. (1966). The problem of universals in language. In J. H. Greenberg (Ed.), *Universals of language* (pp. 1–29). Cambridge: MIT Press.

Holland, J. (1975). *Adaptation in natural and artificial systems*. Ann Arbor: Michigan University Press).

Klopf, A. H., & Gose, E. (1969). An evolutionary pattern recognition network. *IEEE Transactions of Systems Science and Cybernetics, 5*, 247–50.

Moorehead, P. S., & Kaplan, M. M. (Eds.). (1967). *Mathematical challenges to the NeoDarwinian interpretation of evolution*. Philadelphia: The Wistar Institute Press.

Nagel, E. (1932). Measurement. Reprinted In Danto, A. and Morganbesser, S. (1960). *Philosophy of science.* New York: World Publishing Co.

Newell, A. (1980). Physical symbol systems. *Cognitive Science, 4*, 135–83.

Pattee, H. (1972a). The nature of hierarchical controls in living matter. In R. Rosen (Ed.), *Foundations of mathematical biology* (Vol. 1, pp. 1–22). New York: Academic Press.

Pattee, H. (1972b). Physical problems of decision-making constraints. *International Journal of Neuroscience, 3*, 99–106.

Pattee, H. (1982). Cell psychology: An evolutionary approach to the symbol-matter problem. *Cognition and Brain Theory, 5*(4), 325–41.

Polanyi, M. (1968). Life's irreducible structure. *Science, 160*, 1308–12.

Prigogine, J. (1980). *From being to becoming: Time and complexity in the physical sciences*. San Francisco: W.H. Freeman & Co.

Pylyshyn, Z. (1980). Computation and cognition: Issues in the foundations of cognitive science. *The Behavioral and Brain Sciences, 3*, 111–69.

Thom, R. (1975). *Structural stability and morphogenesis*. Reading: W.A. Benjamin.

Turvey, M., & Carello, C. (1981). Cognition: The view from ecological realism. *Cognition, 10*, 313–21.

Turvey, M., & Kugler, P. (1984). An ecological approach to perception and action. In H. T. A. Whiting (Ed.), *Human motor actions: Bernstein reassessed*. Amsterdam: North Holland Publishing Co.

von Neumann, J. (1966). *Theory of self-reproducing automata.* Edited and completed by A. W. Burks. Urbana: University of Illinois Press.

Waddington, C. H. (1968). The basic Ideas of biology. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Prolegomena, Vol. I, pp. 1–32). Edinburgh: Edinburgh University Press.

12 Instabilities and Information in Biological Self-organization

Abstract Two classes of self-organizing systems have received much attention, the *statistically unstable systems* that spontaneously generate new dynamical modes and *information-dependent systems* in which nonstatistical constraints harness the dynamics. Theories of statistically unstable systems are described in the language of physics and physical chemistry, and they depend strongly on the fundamental laws of nature and only weakly on the initial conditions. By contrast, the information-dependent systems are described largely by special initial conditions and constraints, and they depend only weakly, if at all, on the fundamental laws. This results in statistically unstable theories being described by rate-dependent equations, while the information-dependent systems are described by rate-independent (nonintegrable) constraints. It is argued that an adequate theory of biological self-organization requires that these two complementary modes of description be functionally related, since the key process in morphogenesis is the harnessing of cellular dynamics by the informational constraints of the gene. This could arise if the triggering role of fluctuations could be displaced by informational constraints in the control of the dynamical behavior. However, the spontaneous replacement of chance fluctuations by deterministic informational codes is itself a serious problem of self-organization. At present the only approach requires complementary modes of description for the molecular informational constraints and for the macroscopic dynamical behavior that they harness.

What do we expect to learn from a theory of biological self-organization? What types of observables or events do we begin with, and what regularities or laws would we accept as explanations of the self-organizing consequences of these events? Is a theory of biological self-organization fundamentally different from a theory of evolution,

---

Reprinted from *Self-Organizing Systems: The Emergence of Order*, F. E. Yates, Ed. New York: Plenum Press, 1987, pp. 325–338.

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7, DOI 10.1007/978-94-007-5161-3_13, © Springer Science+Business Media Dordrecht 2012

a theory of development, or a theory of thermodynamics; or is self-organization only the result of special types of behavior, such as the singularities found in dynamical systems, or the genetic shuffling found in evolutionary systems? If we survey the many approaches to the problem taken by the other authors of this volume, we might conclude that everyone sees self-organizing behavior in terms of his or her own discipline's theoretical framework. This is to be expected. We all use what we know to better comprehend what we do not know, as Newton used his theory of gravitation better to comprehend God. Nevertheless, even if all disciplines have something to contribute to our comprehension of self-organizing behavior, we still must be able to distinguish this "special type" of behavior from the "normal" events associated with each discipline, if self-organization is to have a distinct meaning.

This chapter focuses on two classes of self-organizing system that have received the most attention: the *statistically unstable systems* and the *information-dependent systems.* The question to be addressed is: What do these two types of systems have to do with each other? Reading the chapters in this book, we find that instability and information are nearly disjoint subjects that are not even discussed in the same language. The chapters on instabilities contain the language of physics or physical chemistry and are particularly sensitive to the basic laws of nature, whereas the chapters focusing on informational concepts contain biological language and are not concerned with physical laws at all. Of course this might simply reflect the differences of the two disciplines of physics and biology. It is also true that authors of papers on instabilities tend to use relatively simple nonliving chemical systems as examples, whereas those writing papers involving information discuss organs as complex as the brain of humans. Again, the differences in approach might simply reflect this enormous difference in levels of organization chosen for study.

These reasons for different approaches are understandable, but ignoring either dynamical instability or symbolic information evades a fundamental aspect of biological self-organization. Although it is true that dynamical theory and symbolic information are not associated in our normal way of thinking, they are epistemologically complementary concepts that are nevertheless both essential for a general theory of biological self-organization. Moreover, instabilities are the most favorable condition of a dynamical physical system for the origin of nondynamical informational constraints, and the evolution of self-organizing strategies at all levels of biology require the complementary interplay of dynamical (rate-dependent) regimes with instabilities and nondynamic (rate-independent, nonintegrable) informational constraints (Pattee 1971). Finally, after discussing these points, I shall comment on the limitations of self-organization theories in terms of epistemological and methodological reductionism.

12.1 The Necessity of Stability

One epistemological requirement of scientific explanation of events is some form of homomorphism between the behavior of the theory as a model or simulation and the behavior of corresponding events or measurements. Or, as Hertz (1894) expressed it, we have an explanation "…when the necessary consequents of the images in thought are always the images of the necessary consequents in the nature of things pictured." This classical epistemological requirement for a correspondence between object and image implies the basic concept of stability that all sufficiently small changes in the images or descriptions of events represent some corresponding, small changes in the events themselves. Again, it is hard to improve on the classical formulations. Poincaré (1952) called a system stable when "small causes produce small effects." Although Newton had no formal methods for testing the stability of his laws of motion, he clearly understood the profound importance of stability for an explanation, since he argued that one of God's primary functions was to maintain the stability of the planetary orbits against the perturbations of other planets. Deterministic laws of motion alone do not ensure stability.

When Laplace proved that Newton's equations were inherently stable, the universe became a great deterministic machine that required no divine intervention once it was created. The only remaining problem, as Laplace (1951) saw it, was the practical impossibility of the human mind acquiring a complete and precise set of initial conditions for all the bodies in the universe. The best he could do was to develop the theory of probability so that we could deal rationally with our remaining ignorance of initial conditions. From this classical, objective epistemology of Newton and Laplace, the idea of self-organization made little sense. Either God intervened locally to keep motions stable, or any unpredictable behavior could be attributed only to our ignorance of details. In either case, there was no apparent desire for additional theories of self-organization or emergent behavior.

12.2 The Necessity of Instability

Logically, any theory of self-organization must accept the precondition of a disorganized subsystem, or a partially disorganized system, since an inherently totally organized system leaves no room for more organization without either redundancy or contradiction. Thus, in the ideal Laplacean universe where every microscopic initial condition is precisely given, we have a totally organized system with an inexorably determined past and future. The well-known escape from this complete determinism, which Laplace recognized, is the condition of human ignorance of initial conditions. This condition is the basis of statistical mechanics; but the most general consequence of this condition was only an increasing disorganization or entropy in the course of time.

The main contribution of Prigogine's school has been to find a new description of how instabilities in these statistical systems may result in entirely new structures from chaotic initial conditions. However, both the ontological and epistemological status of instability remain a fundamental and still controversial problem. One may briefly, and somewhat naively, describe the problem as follows: Our knowledge of the world depends upon our ordering of experience by images, models, descriptions, theories, and so on, which we try to make as clear and unambiguous as possible. To test our theories we predict the consequences from observed initial conditions using the rules of the theory and then look for the corresponding consequences in our later observations. If this correspondence holds over a wide enough range of initial conditions, and if the theory has other coherent logical and esthetic properties, which we find difficult to define, then we may feel that we understand or have described some laws of nature, or some laws of knowledge, depending on our metaphysics.

Now, although we may observe unstable behavior, both directly in events or in our mathematical models of events, we still have a problem in establishing their correspondence by the same criteria we use to establish the correspondence between predictions and measurements of stable behavior. This results from the fact that we recognize events as unstable only because they appear to have no observable deterministic cause, and we recognize descriptions of events as unstable only because the descriptions fail to completely define those events. Poincaré (1952) defined instabilities in the same way that he defined chance events, i.e., as observable events that have no observable cause. In other words, Poincaré's definition of an unstable event implies that it can be described only by a probabilistic model that is logically incompatible with a deterministic model of the same event. The question, then, of whether instabilities are fundamentally the result of chance or determinism in an objective sense or whether they result from the failure of our *descriptions* of events is unresolved and perhaps unresolvable, in any scientific or empirical sense.

In any case we can see that self-organizing behavior must involve something more than stable, deterministic trajectories of classical theories, or the assumption of ignorance of initial conditions, which leads to the stable, deterministic distributions of statistical mechanics. Even though these laws of physics are a foundation for all organization, including self-organization, we recognize in the self-organizing behavior of both nonliving and living systems many entirely new forms and patterns that are not simply the perturbations of stable systems or the probabilistic behavior of unstable systems. The novelty and persistence of emergent forms characteristic of living systems do not fit our definition of either stable or unstable behavior. Therefore, we may expect theories of self-organization also to require complementary subjective or functional modes of description.

Thom's (1975) catastrophe theory of self-organization may at first appear to have little relation to these concepts, since his starting point is pure mathematics. However, Thom's concept of form is also expressed in terms of stability; whatever remains stable under a small perturbation has the same form. Change of form or morphogenesis consequently involves instability. The mathematical concept of "structural stability," as it is called, is technically complex, but can naturally be associated with two complementary modes of description—a control space (subjective) description and a corresponding, state-space (objective) description. The essence of structural stability is that a gradual change in the control-space description induces a corresponding, gradual change in the state space description ("small causes produce small effects"). In the other case, where a gradual change in the control-space description induces a sudden, discontinuous change in the state-space description (bifurcation), there is a change of form, or what Thom calls a catastrophe and we call a form of self-organizing behavior. Of course, many other epistemological interpretations of the formalism are possible since the model, or "method," as Thom calls it, is essentially mathematical. However, it is fundamental to the method that two complementary structures are necessary for developing the concept of catastrophe.

12.3 The Necessity of Complementarity

The modern concept of complementarity is associated with Bohr and the development of quantum theory (Bohr 1928, 1963; Jammer 1974; d'Espagnat, (976), but Bohr believed that the concept was of much more general epistemological significance. Although it was the theory of the electron that forced recognition of complementary modes of description, Bohr felt that complementarity "bears a deep-going analogy to the general difficulty in the formation of human ideas, inherent in the distinction between subject and object." Bohr's definitions of the complementarity principle were not formal or precise, and they have generated much controversy. This is to be expected of any epistemological principle that claims both universal applicability as well as some empirical necessity. It is generally accepted, however, that the principle includes at least two components: first, that to account for or to explain an observed event, two distinct modes of description or representations are necessary, and second, that these two modes of description are incompatible, both in the logical sense that a contradiction would arise if the two descriptions were combined into one formal structure and in the conceptual sense that trying to combine the meanings of both descriptions into one image leads to confusion. Although my concept of complementarity was greatly influenced by reading Bohr and his interpreters, I do not wish to defend or attack his epistemology. I simply find no alternative but to accept multiple, formally incompatible descriptions as a satisfactory explanation of many types of biological events (Pattee 1979).

Perhaps the most fundamental epistemological complementarity arises in our perception of events as either deterministic or chance. The Laplacean ideal of determinism is certainly more than a rational hypothesis about point masses and universal gravitation. One of the most easily observed beliefs of a 5-or 6-year-old child is the child's assumption that every event has a cause or that events could not be capricious (Piaget 1927). Of course, the "causes" that children see are usually animistic or moralistic, but in any case, the idea of determinism is very primitive and does not easily die out in the course of intellectual development. For example, a mature Wigner (1964) characterizes his acceptance of explanation as the feeling that "events could not be otherwise," and most of us are at least emotionally sympathetic with Einstein's belief that "God does not play dice."

The concept of chance comes developmentally with experience, but is never assimilated into our thought with the clarity of the concept of determinism. In fact, gamblers and physicists alike behave as if chance is only determinism disguised by ignorance. The two types of classical theory illustrate this contrast: the so-called microscopic, deterministic descriptions based on the Laplacean ideal of total knowledge and the complementary macroscopic, statistical descriptions based on some predefined ignorance. The former theories are usually pictured as laws of nature that are inexorable in every detail, whereas the latter theories assume some alternative behaviors that are ascribed to chance. The deterministic laws of nature are reversible (time-symmetric) and give predictions that are crucially dependent on knowledge of initial conditions, whereas statistical laws are irreversible and give predictions that are more or less independent of knowledge of initial conditions.

Much of what we call nonliving organization is explained by one or the other of these types of description. Nevertheless, attempts to use both types of theory to explain organization have presented conceptual and formal difficulties since the time of Boltzmann. Prigogine and his collaborators have discovered from their formal attempts to define nonequilibrium entropy that a complementarity principle appears inescapable. The description of deterministic, reversible trajectories is incompatible with a simultaneous description of entropy (Prigogine 1978; Misra 1978). Even without the mathematical formalism, most of us will conceptually agree with Planck (1960) that "it is clear to everybody that there must be an unfathomable gulf between a probability, however small, and an absolute impossibility," i.e., determinism. Therefore, whether one looks at the principle of complementarity as an evasion or as a solution of the problem of determinism and chance, we have very little choice at present but to use complementary models for explaining self-organization in biological systems, in which chance and determinism play such interdependent roles. However, in my biologically oriented epistemology, I am going to suggest that chance is displaced in some optimal sense by informational constraints that efficiently control the higher levels of dynamical behavior.

12.4 The Nature of Dissipative Structures

We have seen that deterministic descriptions are not adequate for explaining self-organizing behavior even at the prebiological levels. The instabilities in deterministic dynamics may be regarded as escape hatches, which in effect leave the behavior of the system undefined in some regions and, hence, subject to unknown or chance events. But chance serves only as an *escape* from classical determinism; it is not a theory of self-organization. The basic contribution of Prigogine was to find an alternative description for the system behavior that exhibits a new structure that is *stable* with respect to the chance events of the previous level. In this way, the instability can serve as a source of chaos in the deterministic, microscopic description and also as a source of new order in the statistical, macroscopic description. However, even though these new modes of behavior effectively introduce a history into physical description, the selection of alternative modes is left to chance.

The physics of this situation is described in terms of far-from-equilibrium thermodynamics, in which instability produces amplification of the fluctuation or chance behavior at the microscopic level and which is then stabilized in the form of new organizations or dissipative structures at the macroscopic level. Although the formalism describing this behavior becomes very complex, it is the epistemological basis of the approach that has the most significance for our theories of self-organization. Classical epistemology, as we said, assigns determinism and objectivity the primary role in theory, with chance and the subjective observer only accepted as unavoidable perturbations in God's universal mechanism. Modern physics, especially quantum theory, forced us to assign a more fundamental role to probability as well as to the role of the observer and the process of measurement. Explanation came to mean not simply reduction of statistical observations to deterministic microscopic events but an effective procedure for correlating our descriptions of observations with our descriptions of laws. Prigogine's epistemological approach to dissipative structures requires this same complementarity between laws and observers. He associates irreversibility, or a direction of time, with the necessary epistemological conditions for observation, thereby achieving a consistent definition of nonequilibrium entropy and the instabilities that allow dissipative structures (Prigogine 1979).

However, this association of instability with the conditions for observation produces an apparent paradox, which is fundamental for theories of self-organization. Recall that the classical concept of instability associates it with chance events—Poincaré's "observable events that have no observable cause." This means that lack of information is associated with chance and instability. Thus, flipping a coin is a chance event only if we do not measure the initial conditions accurately enough. Or, in other words, instability is interpreted classically as a lack of knowledge of the system. But the modem view requires instability as a *condition* for measurement. In this sense, observation and knowledge seem to require a system that is complex enough to display instability in some sense. What is the difference, then, between the instabilities that we associate with loss of knowledge and the instabilities that we associate with the acquisition of knowledge? Or, in terms of theories of self-organization, what is the difference between the organizations that develop from loss of information and the organizations that develop from acquiring new information? One answer is that a loss of information occurs when systems acquire alternative behaviors (bifurcations), while a gain of information occurs when alternatives are reduced (selection).

12.5 The Nature of Symbolic Information

These questions bring us directly back to the fundamental difference between the physicist's approach and the biologist's approach to a theory of self-organization, for despite the new epistemology of physics that gives more weight to probabilistic descriptions and the requirements of observation, there is still an enormous gap between the types of self-organization found in the dissipative structures of macroscopic chemical systems and even the simplest living cells. There is also a serious discrepancy in the idea that any hierarchy of levels of statistical, dissipative structures could ever, by itself, lead to biological self-organization, since the latter is clearly instructed and controlled by individual molecules of nucleic acids and individual enzymes. Statistical mechanics can play no more role in describing these individual molecules than in describing a computer program.

The basic epistemological distinction I wish to make is between the organizations that are constrained by symbolic information and those that develop through chance. Physical theory can be expected to describe self-organizing behavior only insofar as the laws of nature and of statistics are responsible for the behavior; but when the self-organizing behavior is under the constraints of a symbolic information system, then the history of the system dominates our description. The concept of physical laws has meaning only for universal and inexorable regularities, i.e., for systems in which our information about the system can change only with respect to our knowledge of initial conditions. In other words, the form of the laws of nature—even the statistical laws—must be expressed as invariant relative to the information the observer may have about the state of the system. Physical laws do not change in time and have no symbolic memory of past events. On the contrary, observers and symbol systems are characterized by their response to selected past events, which are recorded as memory.

A basic discrepancy, then, between the physicist's and biologist's approach to self-organization is that the physicist's theory recognizes no symbolic restrictions and no historical regularities, whereas the biologist's theory assumes genetic symbol systems with more than three billion years of selected historical structures. The two approaches to self-organizing theories—the instability theories and the information-dependent theories—reflect these two complementary views toward symbols and matter, the instability theories emphasizing fluctuations and ignoring symbolic constraints, and the information-dependent theories ignoring physical laws and emphasizing genetic instructions in their respective formulations. One can find this complementarity sharply distinguished in our two modes of describing computers. On the one hand, the basic electronic gates and memory devices that form the hardware require a description in the language of solid-state physics with no reference to syntax or symbols, whereas the software description is entirely symbolic using programming languages that have no reference to physics. Of course, this same complementarity in our descriptions is required for even the simplest symbolic behavior. The "hardware" description of a pencil has nothing to do with the function of a written message, and the chemical description of a gene has nothing to do with the function of the enzyme whose synthesis it instructs.

We can therefore recognize that the physicist's instability-based concepts and the biologist's information-based concepts of self-organization are also closely related to the two sides of the structure-function complementarity. The concept of function and the concepts of symbol and memory are not a part of physical theory itself. It is significant, however, that when the physicist tries to extend his descriptions to the process of measurement, he cannot avoid the concepts of symbol and function, since measurement is defined as a functional activity that produces symbolic output. For this reason the majority of physical scientists simply ignores the problem of measurement or places it in the biological world, often at the level of the conscious observer. Attempts at unified physical theories of measurement, especially in quantum theory, have not been satisfactory and remain controversial (e.g., see Jammer 1974; d'Espagnat 1976). On the other hand, most biological scientists tacitly assume a classical reductionism they expect will ultimately explain biological activity in terms of physical theory.

When it comes to theories of the brain, cognitive activity, and consciousness, there is not only controversy over what constitutes an explanation, but also basic disagreement on what it is that we are trying to explain. I therefore find it useful to pay more attention to primitive symbol systems where we may expect the matter-symbol relationship to be less intricate, and where the fundamental complementarity of physical instabilities and symbolic information can be more easily explored. My use of the concept of information is strictly limited to semantic information, i.e., to information that is characterized by its meaning, value, or function. Of course, semantic information is not as well defined as is the structural information in communication theory or complexity theory; but the problem of value and function is obviously the central issue in biological organization. My concept of symbol system is essentially the concept of a language—like set of rules (e.g., codes, lexical constraints, and grammars) that are necessary conditions for executing or interpreting symbolic information, instructions, or programs.

12.6 Instabilities and Information

Since the concept of symbolic information is not as well defined as the concept of instability, we need to consider the primitive conditions for symbolic behavior in more detail. The simple concept of a symbol is that it is something that stands for something else by reason of a relation, but it is implicit in this concept that the relationship of symbol to referent is somewhat exceptional. In other words, it is not a physical law. We do not call the electron a symbol for a proton because there is a relationship of attraction between them, nor is the symbol's relationship to its referent the result of statistical laws. We do not call the temperature of a gas a symbol for the velocity distribution of its molecules. In natural languages we say that the symbol-referent relation is a convention. But what does a convention correspond to at the most primitive levels? At the level of the genetic code there is no evidence of a physical or chemical basis for the particular relations between codons and their amino acids. Crick first called this type of relation a frozen accident, and Monod has generalized this biological arbitrariness as a principle of gratuity. To achieve such an arbitrary or conventional aspect of the symbol-referent relation, the physical system in which the symbol vehicles are to exist must exhibit some form of instability. We could also say that a totally deterministic, stable description of the world in which small changes inexorably produce corresponding small effects does not allow the arbitrariness necessary to generate new symbol-referent relations.

Moreover, once created, a symbol system must persist under the same instabilities through which it came to exist, since these instabilities are an inherent property of the underlying physical system. These sound very much like the conditions for Prigogine's dissipative structures, which are created by fluctuations in an unstable thermodynamic regime but which are subsequently stabilized against the same level of fluctuations by the macroscopic coherence of energy flow through the system. However, there is a basic discrepancy between the characteristics of dissipative structures and the characteristics of symbols or informational structures.

Dissipative structures are dynamic, i.e., they depend crucially on the *rate* of matter and energy flow in the system (e.g., reaction and diffusion rates). By contrast, symbol systems exist as *rate-independent* (nonintegrable) constraints. More precisely, the symbol-referent relationship does not depend, within wide limits, on the rate of reading or writing or on the rate of energy or matter flow in the symbol-manipulating hardware. On the other hand, the effect or meaning of symbols functioning as instructions is exerted through the *selective* control of rates. For example, the rate of reading or translating a gene does not affect the determination of which protein is produced. However, the synthesis of the protein as instructed by the gene is accomplished through the selective control, by enzymes, of the rates of individual reactions. At the other extreme of symbol system evolution, we have the example of computers where the rate of reading a program or the rate of computation does not affect the result or what is being computed. On the other hand, the program, as instructions, is actually selectively controlling the rate at which electrons flow in the machine.

A second difference between dissipative structures and symbols is in their size. Dissipative structures occur only when the size of the system exceeds some critical value that is significantly larger than the fluctuating elements that trigger the instability that creates them. This restriction may also be considered as a statistical requirement for a large number of elements to allow stabilization of the new structure. Symbol vehicles, whether bases in nucleic acids, synaptic transmitters, or gate voltages in a computer, are generally not large relative to the size of the organizations they control, nor are symbols essentially statistical in their structure or behavior. A symbol is a localized, discrete structure that triggers an action, usually involving a more complex system than the symbol itself. Also, symbolic inputs are generally amplified in some sense. In other words, symbols act as relatively simple, individual, nondynamical (nonintegrable) constraints on a larger dynamic system.

This suggests that with respect to their relative size, discreteness, nondynamical behavior, and triggering action on larger dynamical systems, symbols act at the same level as the fluctuations that generate dissipative structures. But clearly symbols are completely unlike fluctuations in other respects. Symbol systems are themselves exceptionally stable, and if we are to understand the origin of symbolic behavior at the fluctuation level or molecular level of organization, then we must explain how symbol systems could stabilize themselves without depending on the statistics or averages of macroscopic organization. How can we expect high reliability in molecular information structures that are embedded in a noisy thermal environment?

There are several possible answers to this question, but quantitative results based on theory are still very difficult to produce. The problem was first discussed by Schrödinger (1944), who recognized that the quantum dynamical stationary state in a covalently bonded macromolecule (“aperiodic crystal”) effectively isolated its primary structure from thermal fluctuations. Schrödinger also suggested the analogy of a true “clockworks” at the molecular level, based on quantum dynamical order—and not the statistical order of classical clocks.

London (1961) also speculated that individual enzyme molecules may function in some form of quantum superfluid state that allows the advantage of stationary states along with the possibility of internal motions isolated from thermal fluctuations. In a discussion of the physical basis of coding and reliability in biological macromolecules, I proposed that the correlation of specificity and catalytic rate control in the enzyme required a quantum mechanical non-integrable (rate independent) constraint formally analogous to a measurement process (Pattee 1968). However, no quantum mechanical models of a quantitative relationship of specificity and catalytic power exist at this time.

We should also mention von Neumann's (1966) ideas on the logical requirements for reliable self-replication. Although he gave no proof, or even a precise statement of the problem, von Neumann conjectured that in order to sustain heritable mutations without losing the general self-replicative property, there must be a separate *coded description* of the "self" that is being replicated along with the general translation and synthesis mechanism that reads and executes the description (the "universal constructor").

The most fully developed study of the reliability requirements for hereditary propagation were begun by Eigen and developed into an extensive theory of self-organization at the chemical kinetic level (Eigen and Schuster 1977). Given that some error in information-processing including the template replication of polymer sequences is physically inevitable, Eigen and Schuster show how cooperation between replicating sequences, each with error-restricted information capacity, can lead to hierarchical systems of greater and greater capacity. Their dynamical treatment of this problem requires a special assumption or special boundary condition that I would associate with arbitrary measurement constraints rather than with laws of nature. This in no way weakens their mathematical arguments; however, it bears directly on the question of epistemological reductionism. The special assumption is the existence of specific catalysts that are coordinated to form a primitive code. I want to make it clear that there is nothing wrong with this assumption—indeed, this may actually be the way life began. All I claim is that a code is not reducible to physical laws in any explanatory sense, i.e., without basically revising the concept of explanation. The practical question of origins is simply whether any such functional code had a reasonable probability of occurring by *chance* under primitive earth conditions. The epistemological question is whether the logical concept of codes in general is derivable from only physical laws or whether the concept of code also requires a complementary functional mode of description. In my earlier discussions of the relationship of specific catalysts to the problem of measurement, I concluded that symbolic information can only originate from processes epistemologically equivalent to measurements (Pattee 1968, 1979). Measurement may occur over an enormous range of levels of biological organization, from the specific catalysis of enzymes to natural selection processes that are the ultimate origin of symbolic information. The artificial measurements of physicists fall somewhere in this hierarchy. Syntactical constraints or codes are analogous to the constraints embodied in measuring devices. Without such coherent constraints, neither informational nor measurement processes would have any function or meaning; yet these constraints are arbitrary in the sense that many physically distinguishable syntactical constraints (i.e., different languages and measuring devices) may produce indistinguishable meanings or results. This redundancy or degeneracy is found in all levels of symbol systems from the genetic code to human languages and contributes to the stability and reliability of symbolic information. One is tempted to contrast Poincaré's concept of instability, i.e., distinguishable events for which we find no distinguishable antecedents, with this degeneracy characteristic of symbol vehicles, i.e., distinguishable antecedents for which we may find no distinguishable consequence However, this is too simple a comparison, for even in the most reliable symbol systems some unstable ambiguities in function must exist with respect to perturbations in symbol vehicle structure. I shall return to these epistemological issues in the last section, but my main purpose here is to suggest how dynamics, instabilities, dissipative structures, and symbolic information are related in biological self-organization.

I begin with the hypothesis that the elemental basis for the symbol-referent relationship is the individual specific catalytic polymer in which the folded shape of the molecule is arbitrarily or "gratuitously" coupled to control a specific dynamical rate. This is the only general type of coupling that provides the necessary conditions for a symbol-referent relationship, although it is by no means sufficient. The recognition site or substrate binding site is not related to the catalytic site simply by dynamical or statistical laws. Rather, it requires the particular rate-independent constraints of the folded polymer to determine what molecules are recognized and what bonds are catalyzed. Given this type of specific catalyst, a true code relationship between structure and dynamics is logically possible, along with self-replication and Darwinian natural selection. The question is, can we predict some general organizational consequences of this complementary view of dynamics and information? We expect all stable dynamical behavior to be largely autonomous; that is, since stable dynamics generate no alternatives, there is no need for informational control except in constraining internal boundary conditions. Of course in a stable dynamical regime there still are fluctuations (variations) in these boundary conditions, e.g., modifying protein sequences with or without selection. However, if the limits of dynamical stability are exceeded through excessive competition or new interactions, the informational constraints become dominant in choosing a new stable dynamical structure. This should result in a sudden, large, phenotypic change incommensurate with any structural information measure that could be observed and produce two distinguishable patterns of genotypic and phenotypic change. Under stable phenotypic dynamics there may be observable structural changes in the gene with little corresponding phenotypic change. This would superficially appear as selective neutrality. However, under unstable phenotypic dynamics, a major evolutionary change, like speciation, could result from only minor genetic change. This situation would of course appear to be even more complex, because dynamical instabilities will evolve at all levels of the organizational hierarchy.

A second consequence of this complementary view of self-organization may appear in our approach to the evolution of symbol systems themselves. Since symbolic information at all levels, from nucleic acids to natural languages is so obviously effective for instruction and control of dynamical systems, we usually jump to the conclusion that this capability is an intrinsic and autonomous property of the symbol system alone. From our view of the symbol-matter relationship, however, there is virtually no meaning to symbols outside the context of a complex dynamical organization around which the symbolic constraints have evolved. It is useless to search for the meaning in symbol strings without the complementary knowledge of the dynamic context, especially since the symbolic constraints are most significant near dynamical instabilities.

From this point of view the relationship of the present design of computers and their programming languages represents a bizarre extreme. The hardware is designed to have no stable dynamics at all. The gates and memories are a dense maze of instabilities that require explicit, detailed programs of informational constraints for every state transition. Programming languages therefore have no natural grammars. Machine language is conceptually vacuous, whereas high-level languages that try to mimic natural languages are generally compiled or translated to machine code only at the cost of speed and efficiency. The benefit of this design, of course, is universality of computation. The central epistemological problem of computer simulations of such complex systems as the brain is the ability to distinguish the part of the computer simulating the brain's dynamics from the part simulating the brain's information, since in the universal computer all dynamics are simulated by information. This raises the deeper question of whether or not we can determine if the brain itself uses a dynamical mode for its representations or if all knowledge is restricted to informational constraints.

12.7 Epistemological Limitations

The conclusion that it is useless to search for meaning in symbols without complementary knowledge of the dynamics being constrained by the symbols raises the classical issue of what it would mean to "know" the dynamics. We can mention only briefly here two major schools of thought: that of the information processors who believe that to know the dynamics means representing the dynamics with yet another symbol system (e.g., Newell and Simon 1972) and that of the subjectivists who believe that to know the dynamics is tacit or ineffable, i.e., that it cannot be represented by any symbol system (e.g., Polanyi 1958). In light of these opposing concepts of knowing, we may reconsider the question of whether symbols can be epistemologically reduced to dynamics, i.e., whether codes can be derived from physical laws. To information processors the reduction of symbols to dynamics is not logically possible, since to them, knowing means reducing dynamics to symbols. To a subjectivist the reduction of symbols to dynamics would place knowing entirely in the realm of the ineffable, which, although acceptable epistemologically, is methodologically impotent.

I do not believe that any of these arguments, including my own, are likely to be convincing at the level of the nervous system because of the complexity of the hierarchical organization of both its dynamical and symbolic modes. There is no question that information-processing at many coded symbolic levels goes on in the brain. It is also obvious that much of what goes on in the brain is unconscious and inaccessible to objective analysis. However, it is largely because of these difficulties that studying the relationship of symbol systems to dynamics at the molecular level may uncover useful concepts and organizational principles. At least at this level we can say that the symbolic instructions of the gene go only as far as the primary sequence of the proteins. From there, thus constrained, the dynamical laws take over. To the gene, these dynamics are ineffable.

References

Bohr, N. (1928). The Como lecture, reprinted in *Nature 121,* 580.

Bohr, N. (1963). *On atomic physics and human knowledge*. New York: Wiley-Interscience.

d'Espagnat, B. (1976). *Conceptual foundations quantum mechanics*. New York: Benjamin.

Eigen, M., & Schuster, P. (1977). The hypercycle: A principle of natural self-organization. *Naturwissenschaften, 64*, 541–565; 65:7–41, 341–369.

Hertz, H. (1894). *Die Principien der Mechanik in neuem Zusammenhange dargestellt* (Leipzig). English translation: *The principles of mechanics.* New York: Dover 1956.

Jammer, M. (1974). *The philosophy of quantum mechanics*. New York: Wiley.

Laplace, P. S. (1951). *A philosophical essay on probabilities*. New York: Dover.

London, F. (1961). *Superfluids* (2nd ed., Vol. I, p. 8). New York: Dover.

Misra, B. (1978). Nonequilibrium entropy, Lyapounov variables, and ergodic properties of classical systems. *Proceedings of the National Academy of Sciences of the United States of America, 75*, 1627–1631.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.

Pattee, H. H. (1968). The physical basis of coding and reliability in biological evolution. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Vol. I, pp. 69–93). Edinburgh: University of Edinburgh Press.

Pattee, H. H. (1971). Physical theories of biological coordination. *Quarterly Reviews of Biophysics, 4*, 255–276.

Pattee, H. H. (1979). The complementarity principle and the origin of macromolecular information. *BioSystems, 11*, 217–226.

Piaget, J. (1927). *The child's conception of physical causality*. London: Routledge & Kegan Paul.

Planck, M. (1960). *A survey of physical theory* (p. 64). New York: Dover.

Poincaré, H. (1952). *Science and method.* F. Maitland (trans.). New York: Dover.

Polanyi, M. (1958). *Personal knowledge*. London: Routledge & Kegan Paul.

Prigogine, I. (1978). Time, structure and fluctuations. *Science, 201*, 777–785.

Prigogine. I. (1979). Discussion in P. Buckley and D. Peat (eds.), *A question of physics* (p. 74). Toronto: University of Toronto Press.

Schrödinger, E. (1944). *What is life*. London: Cambridge University Press.

Thom, R. (1975). *Structural stability and morphogenesis*. New York: Benjamin.

von Neumann, J. (1966). *Theory of self-reproducing automata.* Edited and completed by A. W. Burks. Urbana: University of Illinois Press.

Wigner, E. P. (1964). Events, laws and invariance principles. *Science, 145*, 995–999.

13 Evolving Self-reference: Matter, Symbols, and Semantic Closure

Abstract A theory of emergent or open-ended evolution that is consistent with the epistemological foundations of physical theory and the logic of self-reference requires complementary descriptions of the material and symbolic aspects of events. The matter-symbol complementarity is explained in terms of the logic of self-replication, and physical distinction of laws and initial conditions. Physical laws and natural selection are complementary models of events. Physical laws describe those invariant events over which organisms have no control. Evolution by natural selection is a theory of how organisms increase their control over events. A necessary semantic closure relation is defined relating the material and symbolic aspects of organisms capable of open-ended evolution.

13.1 What Is Self-reference?

Self-reference has many meanings. In symbol systems, like logic and language, self-reference may lead to well-known ambiguities and apparent paradoxes as in, “This sentence is false.” In material systems, like molecules and machines, self-reference is not clearly defined but may describe causal loops such as autocatalytic cycles, feedback controls, and oscillators. At the cognitive level, self-reference occurs in introspection and is often considered one aspect of consciousness. I define a specific form of self-reference that applies to a closure relation between both the material and the symbolic aspects of organisms. I argue that this view of self-reference is necessary to understand open-ended evolution, development, and learning at all levels of organization from the origin of life to the cognitive level. This is not an entirely new view, but is an elaboration and integration of ideas from several well-established areas of physics, logic, computation theory, molecular biology, and evolution theory. To state my position as briefly as possible, self-reference that has

---

Reprinted from *Communication and Cognition—Artificial Intelligence*, 12(1–2), 9–28, 1995.

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7,
DOI 10.1007/978-94-007-5161-3_14,

open-ended evolutionary potential is an autonomous closure between the dynamics (physical laws) of the material aspects and the constraints (syntactic rules) of the symbolic aspects of a physical organization. I have called this self-referent relation *semantic closure* (Pattee 1982) because only by virtue of the freely selected symbolic aspects of matter do the law-determined physical aspects of matter become functional (i.e., have survival value, goals, significance, meaning, self-awareness, etc.). Semantic closure requires complementary models of the material and symbolic aspects of the organism. This brief statement requires much more elaboration.

I have emphasized in many papers (e.g., Pattee 1969, 1972, 1982) that the matter-symbol distinction is not only an objective basis for defining life but a necessary condition for open-ended evolution. My reasoning is based not only on biological facts but on the principled epistemic requirements of physical theory. In other words, I require that *models of living systems must be epistemologically consistent with physical and logical principles.* It is well known that replication and evolution depend crucially on how the material behavior of the organism is influenced by symbolic memory. Biologists call this matter-symbol distinction the phenotype and genotype. Computationalists call this the hardware-software distinction. Philosophers elevate this distinction to the brain-mind problem. What is not as well known is that even in the formulation of physical theories a form of matter-symbol distinction is necessary to separate laws and initial conditions. I will explain this further in Sect. 13.4.

The logical necessity of this matter-symbol complementarity was first recognized by von Neumann (1966) in his discussion of self-replicating automata that are capable of creating more and more complicated automata. This is often called emergent evolution. Von Neumann noted that in normal usages matter and symbol are categorically distinct, i.e., neurons generate pulses, but the pulses are not in the same category as neurons; computers generate bits, but bits are not in the same category as computers, measuring devices produce numbers, but numbers are not in the same category as devices, etc. He pointed out that normally the hardware machine designed to output symbols cannot construct another machine, and that a machine designed to construct hardware cannot output a symbol. This was a simple observation about actual machines and the use of natural language, not an ontological or dualistic assertion. Von Neumann also observed that there is a "completely decisive property of complexity," a threshold below which organizations degenerate and above which open-ended complication or emergent evolution is possible. Using a loose analogy with universal computation, he proposed that to reach this threshold requires a universal construction machine that can output any particular material machine according to a symbolic *description* of the machine. Self-replication would then be logically possible if the universal constructor is provided with its own description as well as means of copying and transmitting this description to the newly constructed machine.

As in the case of the universal computing machine, to avoid the ambiguities of self-reference, logic requires the categorical distinction between a machine and a description of a machine. This logic does not differ if the machine is a material machine or only a formal machine. To avoid the ambiguities of self-reference requires two logical types or categories. This is the logical basis of the matter-symbol distinction. It is significant that his so-called kinetic model required primitive parts with both symbolic functions (i.e., logic functions) and material functions (e.g., cutting, moving, etc.). I will discuss this argument in Sect. 13.9. Von Neumann made no suggestion as to how these symbolic and material functions could have originated. He felt, "That they should occur in the world at all is a miracle of the first magnitude." This is the origin of life problem.

13.2 What Is Matter?

For my argument here, I will mean by matter and energy those aspects of our experience that are normally associated with physical laws. These laws describe those events that are as independent of the observer as possible, i.e., independent of initial conditions. The laws themselves are moot until we provide the initial conditions by a process of measurement. Laws and measurements are necessarily distinct categories. Laws do not make measurements, individuals make measurements. Measurement is an intentional act that has local significance and hence involves symbolic aspects usually in the form of a numerical record. This is the physical basis of the matter-symbol distinction. I elaborate on this in Sect. 13.5. This well-established distinction between the physical and symbolic aspects matter we have no trouble recognizing in practice. Whether one is a material reductionist or a formalist, in practice we rarely have difficulty distinguishing our descriptions of matter using physical laws and our descriptions of symbols using syntactical rules and programs. Also, we all know the difference between formulating theories, constructing instruments, making measurements, and computing.

The difficulty begins when we try to describe how these complementary material and symbolic aspects are related. Traditional philosophy sees this relation as the problem of reference, or how symbols come to stand for material structures (e.g., Whitehead 1927; Cassirer 1957; Harnad 1990). I have always found the complementary question of how material structures ever came to be symbolic much more fundamental. From the origin of life and evolutionary perspective the most difficult problem is how material structures following physical laws with no function or significance were gradually harnessed by syntactical rules to provide function and significance as symbols (e.g., Pattee 1969). I will not say much more about the origin problem here. For several reasons, one of which is its difficulty, the origin of symbols is not considered one of the central problems in any area of philosophy or science. Another reason is that for most scientific models it is not necessary to know the nature or origin of symbols. Natural language, logic, mathematical symbol systems, and computers are most commonly treated simply as well-developed tools, and for most models there is no need to ask how they originated.

13.3 Material and Symbol Complementarity—Neither Reductionism nor Dualism

The biologist, largely for historical reasons, often considers dualism as the only alternative to reductionism. This is not my view. I am a physical reductionist in the sense that I believe all symbolic behavior must have a material embodiment, following physical laws, that correlates with this behavior. For example, the chemical structure and reactions of DNA must correlate with its function as the record of genetic instructions, the chemical structure and reactions of a photosensitive emulsion must correlate with its function as a record of a measurement process, and the biochemistry of neurons must correlate with all perception and thought. Material reductions are certainly one necessary type of model for understanding symbol systems.

However, I am not a reductionist in the sense of those who claim that symbols are "nothing but" matter. "Nothing but" implies that the only model that is required to understand symbols is a complete materialist or physical law model. Reductionists are generally happy when they have discovered the material correlates of higher level behavior. My position is that no complete physical description of these material structures, although correct in all details, will tell us all we need to know about their symbolic function. Briefly, this is because symbol function, like all biological function, is not an intrinsic or law-based property of the material symbol vehicles but a selective survival property of the populations of individuals that use the symbols for material construction and control in a particular environment. In other words, one exclusive material reductionist model is not adequate to describe function or significance. An alternative, complementary model is necessary. I will elaborate on this in Sect. 13.4.

I want to emphasize that my position did not originate from this metaphysical view, but from the way physicists and biologists actually formulate their models of the world. Therefore, rather than trying to clarify the thorny issues of reductionisms vs. dualisms that historically have been scientifically sterile, I will only elaborate on the well-established value of using complementary models without entering into the undecidable metaphysical issue of which model represents reality. Complementary models are well-known in physics. Particle and wave, microscopic and macroscopic, deterministic and stochastic, coarse and fine grained, reversible and irreversible models are necessary for fully understanding any complex system. Rosen (1977) has usefully defined a measure of system complexity by the number of models that we require to adequately understand its behavior.

13.4 What Is Measurement?

In my opinion, the measurement process in physics is the most convincing and fundamental example of the necessity of complementary models with semantic closure. On the one hand, it is possible to describe a measuring device in its material detail, and this may be necessary in its design and construction. On the other hand, if the measuring device is to perform its function (i.e., produce a symbolic record) these details must be selectively ignored. This is not a metaphysical position but arises from the pragmatic fact that to obtain a meaningful result *we must be able to measure something without having to measuring everything.* This means that to function, the number of material degrees of freedom in the measuring device must be reduced to the few *semantically relevant* symbolic degrees of freedom of the result. Without such a classification process we have a divergent infinite regress of measurements, as von Neumann (1955) pointed out.

This concept of measurement generalizes to all interaction of organisms with their environment that require classification for survival. The distinction between the material and symbolic behavior is very sharp in physical theory for the principled reasons I will explain further in the Sect. 13.5. However, in primitive organisms matter and symbol are not as easy to disentangle. This is the case with all structure-function relations in organisms. As external modelers we need to know the detailed chemical structure of DNA to understand, and perhaps to design, the chemical correlates of its function, but to perform its semantic function in the cell only the cell's classification of the base sequence is relevant to the synthesis of proteins. Just as in the case of a measuring device, there is a great reduction from the many degrees of freedom of the material codon to the few bits of semantic information it actually conveys as a result. Similarly, at the protein level any external structural model of the material folding requires enormous detail and computational power, but folding in the cell is a physical process that requires no description or instruction beyond synthesizing the linear sequence of amino acids. In all cases, from our modeling point of view we cannot ignore multiple-level descriptions when we need to relate structure to function. Similarly, but more objectively, some of the cell's behavior, like reading base sequences, is symbolic, but most of its behavior, like protein folding, is not. That is, DNA symbolically describes only the linear sequence of amino acids, while physical laws take care of folding, self-assembly, and catalysis.

As in the case of measurement, in order to have any useful function, *genes must be able to symbolize something without symbolizing everything.* Otherwise genetic instructions would never end. Without simplification, heritable symbols would suffer the same infinite regress as measurement symbols. Therefore to allow open-ended increase of material complexity while maintaining heritability requires simplification of description. In conventional language, symbolizing something without symbolizing everything is called classification. Consequently organizations with the potential for emergent evolution, above von Neumann's threshold of complication, must perform *autonomous* classifications.

13.5 What Is a Symbol?

Symbols are difficult to define in any simple way because symbols are functional, and function cannot be ascribed to local structures in isolation. The concept of symbol, like the more general concept of function, has no intrinsic meaning outside the context of an entire symbol system as well as the material organization that constructs (writes) and interprets (reads) the symbol for a specific function such as classification, control, construction, communication, decision-making, and model-building (e.g., Pattee 1969). All these activities can be identified as functions only in specific contexts from local goals of individuals to the global survival of species. The symbol vehicle is only a small material structure in a large self-referent organization, but the symbol function is the essential part of the organization's survival and evolution. This autonomous structure-function self-referent organization is what is entailed by my term *semantic closure*.

For this discussion I could alternatively describe a symbol as a relatively simple material structure that, while correctly describable by all normal physical laws, has *significance or semantic function* that is not describable by these laws. The reason that laws cannot describe symbol function, or any function, is because we specifically restrict physical laws to describe only those properties of matter that are, by principles of invariance and symmetry, as independent of observers and individual measurements as possible, as I stated in Sect. 13.2. This is necessary to achieve the characteristic universality of laws. Symbols, by contrast, are generated with few physical restrictions but are eventually selected for their contribution to the survival of individual units in a local environment. In other words, only those universal and intrinsic aspects of matter that have *no* significance for individuals are described by laws, while only those context-dependent, selective aspects of matter that have significance for individuals in a local environment are described as symbols (Pattee 1982). Of course, all symbols require material vehicles that obey all the laws, but symbolic function requires another model. These are complementary models, not dualism.

To understand why physical theory cannot treat symbols as nothing but matter described by laws one must first understand that the present concept of physical law makes sense only if we divide experience into things that don't change and things that do change. This distinction is one of the defining characteristics of laws and initial conditions (Wigner 1964). Furthermore, it is only because of this independence of material and symbolic aspects that physical laws can be modeled with the minimum ambiguity between the boundary of the lawful world and the formal model. In other words, it is the independence of symbolic aspects from material aspects that allows a clear, fundamental separation of laws and initial conditions. There is no way to give much meaning to symmetry, invariance and conservation principles without a sharp separation of laws and initial conditions (e.g., Houtappel et al. 1965). In physics, the act of measurement of initial conditions is the only contact of the symbolic model with the material world. Laws are moot until provided with specific initial conditions by measurements. Therefore symbols must be viewed as belonging to the general category of initial conditions, which also includes boundary conditions and constraints. Ordinary initial conditions are without regularity, but symbols are special collections of constraints that allow us to describe symbolic behavior by rules.

Another explanation of why symbolic behavior cannot be described by laws is that laws are invented to be complete and inexorable. Therefore, one cannot amend or adjust lawful behavior itself. Laws leave no alternatives. The only meaning we can attach to a choice of alternatives in a system described by deterministic laws is through measurement and control of initial conditions. Writing symbols is a time-dependent dynamic activity that leaves a time-independent structure or record. The mathematician Emil Post (1965) described the writing of symbols as, “Activity in time [that] is frozen into spatial properties.” Symbols are read when these structures re-enter the dynamics of laws as constraints (Pattee 1972). Any highly evolved formal symbol system may be viewed as a particularly versatile collection of initial conditions or constraints, often stored in a memory, producing significant or functional behavior that is usefully described by locally selected rules rather than by physical laws. This means that rules for manipulations of the material symbol vehicles are as independent of mass, energy, dynamical time, and rates as possible. The genetic code, natural language, logics, formal mathematics, and computer programming languages are the best known examples of such symbol systems. As I have emphasized, all symbol systems must have material embodiments that obey physical laws. But for the reasons just stated, the lawful material description of symbols, even though correct in all details, can reveal no significance.

13.6 What Is a Symbolic Model?

Any model must in some sense have similar behavior to what it models. In the symbol-based formal models that are the established format for physical theories, similar behavior is a metaphor established by a parallelism between a few selected aspects of behavior of the object, ascribed to inexorable laws, and a few selected aspects of behavior of the symbols, determined by our local mathematical or computational rules. Because the material vehicles of symbols are physically arbitrary (i.e., energy degenerate) structures and their rules based on boundary conditions and not derived from laws, it is a characteristic of symbolic models that outside of these few selected parallel aspects there is generally no other similarity between the material system and the symbolic model. All we can expect from symbolic models is that a few specific aspects of our models and a few specific aspects of the object have similar or predictable behavior.

Hertz (1894) first stated clearly the relation between matter and symbols in a model: “We form for ourselves images or symbols of external objects; and the form which we give them is such that the logically necessary (denknotwendigen) consequents of the images in thought are always the images of the necessary natural (naturnotwendigen) consequents of the thing pictured.” Then, to emphasize the limited domain of reference of formal symbolic models, he added, “For our purpose, it is not necessary that they [the symbols] be in conformity with the things [the matter] in any other respect whatever. As a matter of fact, we do not know, nor have we any means of knowing, whether our conception of things are in conformity with them in any other than this one fundamental respect” It is significant that by substituting “instructions” for Hertz’s first two usages of “images,” and “constructions” for the last usage, we have a concise description of the function of the genetic code.

13.7 How to Evade the Matter-Symbol Problem

Each scientific culture has its own reasons for ignoring the matter-symbol problem. Physics, with its sharp categorical distinction between matter and symbol, does not normally require a theory of symbols even though theories expressed in mathematical symbol systems play the primary role in physics. Also, physicists study material systems that in most cases do not themselves contain intrinsic symbolic activities and functions. In the case of measurement of initial conditions, that is, the mapping of matter to symbols, the measurement is treated as a primitive process for which, fortunately, a theory of symbols is not necessary for useful symbolic results. However, in well-known thought experiments where it is necessary to specify objectively when a measurement is completed, such as Maxwell's demon and Gibb's paradox, the matter-symbol problem is unavoidable (e.g., Leff and Rex 1990); and in quantum theory, where the measurement process enters the formalism of the laws, the interpretation of measurement remains largely inscrutable (e.g., Wheeler and Zurek 1983). In any case, neither the evolutionary origin of symbols nor of measurement processes is considered a dominant issue in physics.

Because all organisms depend on intrinsic symbolic controls and the origin of life requires a symbolic genetic code as a crucial step, biologists should be much more interested in the matter-symbol problem. However this is not the case. Most biologists are material reductionists, and the discovery of the material structures that correlate with the symbolic activity and function is the only level of explanation they are looking for. Consequently, experimental or material discoveries, not theory, play the primary role in biology. For example, the biologist finds the chemical structure of DNA and the molecular basis of coding a satisfactory description and feels that this fully explains the gene's symbolic behavior. This material reductionism is even extended to cognitive activity where discovering the material neural correlates of thought would be considered by many as a satisfactory reduction of conscious behavior (e.g., Crick 1993; Hopfield 1994).

Philosophers have traditionally focused on the higher level mind-body problem, but they have also found metaphysical stances that effectively minimize the matter-symbol problem, such as idealism, dualism, material reductionism, functionalism, and the newest and most effective of all, computationalism. Besides the traditional cultures of philosophy, physics and biology, a fourth computer-based culture comprising the classical field of artificial intelligence (e.g., Newell 1980; Pylyshyn 1984) and the more recent field of artificial life (e.g., Langton 1988) has adopted the programmable computer as a universal symbolic model. This culture explicitly disregards material embodiments of either the computer or what it is modeling. Both artificial intelligence and artificial life evade the matter-symbol problem by accepting a functionalist or the stronger computationalist view of models. Like classical physicalism, functionalism and computationalism also make a sharp categorical distinction between matter and symbol, but they focus only on the symbolic category. Functionalists argue from the half-truth that because there are innumerable possible material embodiments of any given symbol function, the relation of symbols to matter is gratuitous or arbitrary. They believe that the particular facts of biochemistry, neuroanatomy, and neurophysiology represent only one possible material embodiment of biological and mental behavior, and that the computer can, in principle, equally well represent another embodiment. In other words, functionalists consider the particular material embodiment of the symbolic activity as unimportant.

The requirements for functionalist models may sound like they are based on the same classical principles as are the models of physical systems, namely, that selected aspects of the computer model's behavior must parallel selected aspects of the organism's behavior. The computationalist requirements are similar, only the word function now takes a formal symbolic meaning. The main requirement for the computationalist's model is that it computes at least one of the same functions as the object being modeled (e.g., Dietrich 1994). In spite of this apparent similarity of physical and computer models there is a fundamentally different view of the role of measurement and consequently of the matter-symbol relation. Physicists view measurement as the only empirical contact with the world. Therefore their observables are precisely defined, relatively simple, and accurately measurable. In physics, enormous effort and by far the largest amounts of time and resources are spent on designing and constructing measuring devices and on actually performing measurements.

Functionalist and computationalist modeling organism and brains have a much more difficult problem defining observables. Most of their time and resources are spent programming and running computers, and insofar as they use observables, they can seldom define them precisely enough to measure objectively. Artificial intelligence typically models complex cognitive activities such as problem-solving, pattern recognition, or types of thinking. Artificial life typically models activities such as self-replication, adaptation, and emergence. These are not simple enough observables to be precisely defined or measured. Their symbol manipulation is precise enough, but their symbol grounding is vague. Naturally this leads to undecidable arguments, such as where symbols and meaning begin in computers (e.g., Searle's Chinese room), and whether computers are alive. I will suggest other inadequacies of functionalism in Sect. 13.10. Computationalists must also make the gratuitous assumption that all matter is computing, that is, they assume every material thing is computing something if we choose to interpret it as computing. Such a subjective view evades the matter-symbol problem completely (Pattee 1990).

13.8 Self-organization Approaches

Many scientists have taken the reasonable strategy of treating the matter-symbol distinction as originating at some late stage of a general process of spontaneously increasing complexity of material systems. This type of model has often been called self-organization. The older literature includes discussions of physical systems that are described simply, but spontaneously grow in complexity (e.g., Yovits and Cameron 1960; Yovits et al. 1962). The few general theories of self-organization were not thoroughly developed at the time (Simon 1962; Burgers 1963; Pattee 1969;

Kauffman 1969). In the 1970s new types of order production were discovered in nonequilibrium thermodynamical and non-linear dynamical models (e.g., Glansdorff and Prigogine 1971; Haken 1977; Nicolis and Prigogine 1977). More recent work on self-organization is collected in Yates (1987). Some of these developments in thermodynamics have lead to speculations about possible organizing principles that modify the traditional neo-Darwinian model (e.g., Weber et al. 1988). These models and theories of organization are generally applied only to prebiotic or at least pre-symbolic matter, and therefore do not address the matter-symbol relation.

Currently, with the discovery of unexpected richness in nonlinear dynamics, self-organization is now usually included in the new field called the science of complexity (e.g., Stein 1988; Nicolis and Prigogine 1989; Zurek 1990; Stein and Nadel 1990; Waldrop 1992; Kauffman 1993). Its potential arises from many sources that include mathematicians, physicists, and computer and cognitive scientists, each with characteristic but overlapping approaches, e.g., nonlinear dynamics, chaos, cellular automata, non-equilibrium thermodynamics, statistical mechanics, solid-state physics, connectionist machines, artificial neural nets, etc. Even in this new field of complexity theory the origin of symbols is seldom seen as an issue, and most of the computational models in this area do not make any clear distinction between law-based material behavior and rule-based symbol behavior. At least there is no consensus. The field includes some physicists who believe that all our models of physics are limited by the symbolic output of measurements (e.g., Wheeler 1990), and some computationalist who believe that all lawful material processes are computations (e.g., Toffoli 1982). The so-called strong artificial intelligence and strong artificial life modelers believe that particular material embodiments are irrelevant, and consequently that a close enough computer simulation becomes a form of realization of what is modeled (e.g., Langton 1988). Of course if one believes that everything is a computation, or that by improving simulations they will eventually become realizations then one sees no matter-symbol problem (Pattee 1988). At the other extreme there are physical reductionists who see symbols only as an illusion, like phlogiston and the ether, that will be unnecessary when an adequate material description of symbolic behavior is found (e.g., Churchland 1981; Crick 1993).

13.9 The Function of Symbols in Evolution and Cognition

Knowing how protein synthesis works we might conclude that construction was the first function of symbols. However, construction requires the classification and control of parts. Also, construction would be of no evolutionary value unless there was hereditary transmission. This certainly requires communication. In other words, at the primitive levels none of these functions can be isolated as primary nor even objectively distinguished from each other. This is one reason that the origin of symbols and life is such a difficult problem.

At the cognitive level, symbols allow our own subjective sense impressions to be compared with another's and thereby endowed with some degree of objectivity.

By objectivity, here, I mean only the ability of different observers to reach a principled agreement by communicating what they observe. For example, if I see a green light there is no way I can tell another viewer what I see without symbols. Even telling you in natural language that I see “green” is no assurance that your experience of “green” is the same as mine. It is only by a measurement process that abstracts complex perceptions to simpler, more universal, symbolic classifications that agreement, and hence communicable objectivity, is possible among populations of individuals. This is Born’s (1965) fundamental explanation of why mathematical models are essential for representing physical theory. It essentially defines one necessary condition for objectivity in physics. However, the principle is more general than that. The same universal communicable classification is also an essential function of all heritable symbols in populations of organisms capable of evolution by natural selection. The hereditary transmission requirement in evolution is fundamentally a communication problem.

Why does a material structure need symbols to communicate or transfer its structure to other matter? The first reason, mentioned by von Neumann, is that any universal constructor that could assemble its material parts would function more efficiently if it replicated from a symbolic *description* of its material parts rather than replication by material self-inspection of its parts. I have not found a strong physical or logical support for this efficiency argument, although it sounds plausible. In any case, as any evolutionist would point out, replication by symbolic description must be superior to material self-inspection because it survives. The second reason, indirectly suggested by von Neumann, is that new descriptions, being simpler, are more likely to arise than corresponding new material constructions. This is also not a general physical or logical argument, but again, it is plausible from our knowledge of molecular genetics and evolution.

Consider the various physical processes that a simple material structure might use to self-replicate. Assume the structure consists of parts that will self-assemble if brought close enough together. How can the structure bring together the correct parts? Given a reservoir of parts, the simplest way is to have every part of the structure individually heritable, that is, to have each of its parts capable of selecting a similar part from the reservoir. In principle, specially folded macromolecules could do this by template or specific binding, a kind of crystallization with many parts. However, this hereditary process does not have open-ended evolutionary potential, because first, all mutant parts must also have this intrinsic hereditary property. In other words, for this process to achieve open-ended evolution we must assume that the universal heredity property is a rather general intrinsic property of macromolecules. This is not the case. Second, template-identified material structures are limited to the outside parts, just as is crystal growth. The only known way out of the first limitation is to use special adaptors that are universal, that is, a small set of adaptors that can bind any number of correct parts as well as mutant parts. The only known way out of the second limitation is by unfolding to get at the inside parts. We know that molecular adaptors and folding are general strategies of all cells, but the actual implementation is even more complex.

Von Neumann did not discuss the need for adaptors, but it is clear from his solution that he saw the self-referent logical difficulties in postulating universal adaptors that could adapt to themselves. As we outlined in Sect. 13.1, his solution was based on the logical requirements for universality in a formal Turing machine applied to a universal material constructor machine. Generally, a universal machine is one particular adaptor machine that can be instructed to mimic the behavior of any machine. The logical point is that to mimic another machine there must be two categorically distinct levels of instruction. The machine must know when to mimic and when not to mimic. In other words, it must distinguish the virtual machine from the real machine. It is important to understand that universality applies only to the domain of possible symbolic *descriptions* of machines. In other words, mutations may be acceptable in the descriptions, but mutations are not likely to be acceptable in the universal machine itself. This is the case with the genetic code which is the universal reader for an enormous open-ended variety of *descriptions* of proteins. However, the material parts and code itself are essentially the same for all organisms.

13.10 The Role of Matter in Evolution and Cognition

Why are particular material embodiments or hardware important for open-ended evolution if this logic can be satisfied in a symbol system like a computer? Von Neumann himself switched from his kinetic model that recognized the matter-symbol distinction to a formal cellular automaton model that did not. However, he warned that, "By axiomatizing automata in this manner, one has thrown half the problem out the window and it may be the more important half." There is no doubt that programmable computers can simulate many important aspects of life, evolution, and cognitive activity. This has been clearly demonstrated by a vast number of programs. The stronger claims of artificial intelligence and artificial life that a computer can *realize* thought and life are not empirically, or even logically, decidable issues because they hinge entirely on the degree of abstraction one is willing to accept as a realization. If we could agree to define life and thought abstractly so as to leave out enough material aspects then obviously, by definition, a live, thinking computer is possible. Similarly, if we could agree to define the concept of computer broadly enough to include enough material aspects then, by definition, everything may be called a computation. I do not see much value in pursuing this type of undecidable issue. In any case, it is a fact that exclusive models of either symbolic or material aspects of life have not yet answered the functional and semantic issues to everyone's satisfaction.

One inexorable aspect of physical systems that formalists often ignore, or view completely differently from physicists and biologists, is noise. Noise is not only inevitable in all measurements, but is essential for evolution. Computer hardware and neurons are also noisy, but formal models do not recognize noise. By good design of symbol systems and their hardware noise can usually be ignored for the purposes of symbol manipulation. One of the proposed challenges to the Turing Test for assessing whether a machine can think is to see if the machine makes mistakes. Of course, to pass this test one can introduce random error in the program. A more effective and more difficult test is to continually introduce errors in both the challenger and the machine and compare their learning or evolving behaviors.

Formalist and functionalist argue that since different material hardwares can compute the identical function, the computation is independent of is material implementation. This is true only in the ideal case of noiseless, error-free symbol manipulation. However, it is easy to see that two computations that are formally equivalent, i.e., that compute the same function, will generally respond to error in entirely different ways. This occurs even within formal systems. As a simple example, identical bit strings can be generated by the rewrite productions $0 \rightarrow 1$ and $1 \rightarrow 01$ starting with 0, and the recursion $S_{n+1} \rightarrow S_{n-1} \circ S_n$ starting with $S_0 = 0$ and $S_1 = 1$. However, completely different strings will be generated following an error in any bit at any place in the string. This problem only gets worse if the material structures implementing the computation also have noise. Such mutation tests can of course be used to discover if two formally equivalent computations are implemented with different hardware, architectures, or programs. For example, a connectionist machine, cellular automaton, or any number of machines may be formally equivalent, but it is highly unlikely that their response to noise in hardware or software will be similar or even related in any predictable way.

This is clearly a very general type of test, because, in fact, when coupled with heritable memory and natural selection, it is the basis of evolution itself. The functionalist's position that the same function can be realized by many material structures should be countered by three additional physical and biological facts. First, the same material structure can perform different functions, since function is not intrinsic to any structure. Second, the domain and quality of potential functions of a given material structure will depend on details of that material structure. That is, two different material structures will not have the identical domains of potentially evolvable functions. Generally, different material structures will evolve differently, even though at one time they may have both had the same function. Third, effective evolutionary search depends on how the space of symbolic description maps to material functions. The 3-dimensional folding of proteins is not related to their 1-dimensional genetic description only by symbolic rules, but depends crucially on material structures and physical laws (e.g., Conrad 1990).

The role of the material structure that implements cognitive activity in brains is not nearly as well understood as its role in the evolution of organisms. The functionalists and computationalists still apply the same argument that symbolic behavior of brains is not dependent on the material implementation. Again, this is true by definition if cognitive behavior is abstracted far enough from its material-dependent precursors, such as sensorimotor controls. However, the three facts just stated above also apply to the evolution of brains and to learning processes, so it is not likely that the successful evolution of rapid and accurate classification processes, such as complex visual pattern recognition, had no dependence on the material structures through which such functions evolved.

13.11 Conclusions

As I indicated in Sect. 13.1, I require that theories of life be epistemologically consistent not only with logic but with fundamental physical principles. The most fundamental epistemological classification is between things that do not change and things that change. In physics this principle is used to define laws and initial conditions. This implies a self-referent impotency principle that unchanging events cannot completely describe changing events. That is, laws cannot completely describe measurements. More precisely, the classification function of measurement cannot be derived from laws. Otherwise, the laws could derive their own initial conditions by computation. The corresponding self-referent impotency in formal systems is that they cannot prove their own consistency, let alone assign a truth value to their own axioms. This implies that formal symbol systems also cannot make measurements. Symbolic computation can never realize measurement.

Physical laws and natural selection are complementary models of events. Physical laws describe those events over which organisms have no control. Evolution by natural selection is a theory of how organisms increase their control over events. By natural selection I mean the neo-Darwinian process of biasing the relative survival rates of population distributions grown by heritable variations of their symbolic instructions. The biasing is done at many levels of organization (e.g., Sober 1984). We can define non-selective self-organization as order produced by present or future physical laws in systems unconstrained by symbolic instructions.

Kauffman (1993) in his exploration of non-selective ordering processes points out that no established field of study incorporates the non-selective physical order into evolution theory. To some extent this may be another case of cultural bias in scientific models inherited from the classical physicist's categorical distinction of matter and symbol. Perhaps it is also because until recently there has been a lack of specific theories of physical self-organization that appeared to be relevant to biological organisms. This is no longer the case. As I mentioned in Sect. 13.8, there have been many recent discoveries of complex physical systems that exhibit emergent order that to many appear lifelike. However, the matter-symbol distinction is rarely addressed in these studies. Only theories of the origin of the genetic code appear directly relevant to the matter-symbol distinction (e.g., Bedian 1982). As in the case of artificial intelligence, computational models of emergent evolution while stimulating new interest in the classical matter-symbol problem, have rarely addressed the physical basis for the distinction or how matter and symbol are related by measurement.

For all these reasons, I find that a productive approach to the theories of life, evolution, and cognition must focus on the complementary contributions of non-selective law-based material self-organization and natural selection-based symbolic organization. To some degree the nature of this complementary relation is an empirically decidable issue. However, it is also a foundational issue. The semantic closure of dynamical laws and symbolic constraints is a necessary epistemological condition for information, knowledge, models, and theories at all levels of evolution from the genes to the brain.

References

Bedian, V. (1982). The possible role of assignment catalysts in the origin of the genetic code. *Origins of Life, 12*, 181.

Born, M. (1965). Symbol and reality. *Universitas, 7*, 337–353. Reprinted in Born, M. *Physics in my generation* (pp. 132–146). New York: Springer-Verlag.

Burgers, J. M. (1963). On the emergence of patterns of order. *Bulletin of the American Mathematical Society, 69*, 1–25.

Cassirer, E. (1957). *The philosophy of symbolic forms* (The phenomena of knowledge, Vol. 3). New Haven: Yale University Press.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy, 78*(2), 67–90.

Conrad, M. (1990). The geometry of evolution. *BioSystems, 24*, 61–81.

Crick, F. (1993). *The astonishing hypothesis*. New York: Scribner's Sons.

Dietrich, E. (Ed.). (1994). *Thinking computers and virtual persons*. New York: Academic.

Glansdorff, P., & Prigogine, I. (1971). *Thermodynamics of structure, stability, and fluctuations*. London: Wiley.

Haken, H. (1977). *Synergetics*. Berlin: Springer.

Harnad, S. (1990). The symbol grounding problem. *Physica D, 42*, 335–346.

Hertz, H. (1894). *The principles of mechanics* (pp. 1–2). New York: Dover [orig. German ed., *Prinzipien Mechanik*, 1894] Quoted in Weyl, H. (1942). *Philosophy of mathematics and natural science* (p. 162). Princeton: Princeton University Press.

Hopfield, J. J. (1994). Physics, computing, and why biology looks so different. *Journal of Theoretical Biology, 171*, 53–60.

Houtappel, R. M. F., Van Dam, H., & Wigner, E. P. (1965). The conceptual basis and use of the geometric invariance principles. *Reviews of Modern Physics, 37*, 595–632.

Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly connected nets. *Journal of Theoretical Biology, 22*, 437.

Kauffman, S. A. (1993). *The origins of order*. New York: Oxford University Press.

Langton, C. (1988). *Artificial life*. Redwood City: Addison-Wesley.

Leff, H. S., & Rex, A. F. (Eds.). (1990). *Maxwell's demon, entropy, information, computing*. Princeton: Princeton University Press.

Newell, A. (1980). Physical symbol systems. *Cognitive Science, 4*, 135–183.

Nicolis, G., & Prigogine, I. (1977). *Self-organization in non-equilibrium systems*. New York: Wiley.

Nicolis, G., & Prigogine, I. (1989). *Exploring complexity*. New York: Freeman.

Pattee, H. H. (1969). How does a molecule become a message? *Developmental Biology Supplement, 3*, 1–16.

Pattee, H. H. (1972). Laws, constraints, symbols, and languages. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Vol. 4, pp. 248–258). Edinburgh: Edinburgh University Press.

Pattee, H. H. (1982). Cell psychology: An evolutionary approach to the symbol-matter problem. *Cognition and Brain Theory, 5*(4), 325–341.

Pattee, H. H. (1988). Simulations, realizations, and theories of life. In C. Langton (Ed.), *Artificial life* (pp. 63–77). Redwood City: Addison-Wesley.

Pattee, H. H. (1990). Response to E. Dietrich's "Computationalism". *Social Epistemology, 4*(2), 176–181.

Post, E. (1965). Quoted from M. Davis (Ed.), *The undecidable.* Hewlett: Rowen Press.

Pylyshyn, Z. (1984). *Cognition and computation*. Cambridge, MA: MIT Press.

Rosen, R. (1977). Complexity and system description. In W. E. Harnett (Ed.), *Systems, approaches, theories*. Boston: Reidel.

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

Sober, E. (1984). *The nature of selection*. Cambridge, MA: MIT Press.

Stein, D. L. (Ed.). (1988). *Lectures in the sciences of complexity*. Redwood City: Addison-Wesley.

Stein, D. L., & Nadel, L. (Eds.). (1990). *Lectures in complex systems*. Redwood City: Addison-Wesley.

Toffoli, T. (1982). Physics and computation. *International Journal of Theoretical Physics, 21*, 165–175.

von Neumann, J. (1955). *Mathematical foundations of quantum mechanics*. Princeton: Princeton University Press. Chapter VI.

von Neumann, J. (1966). In A. Burks (Ed.), *The theory of self-reproducing automata*. Urbana: University of Illinois Press.

Waldrop, M. M. (1992). *Complexity*. New York: Simon & Schuster.

Weber, B. H., Depew, D. J., & Smith, J. D. (1988). *Entropy, information, and evolution*. Cambridge, MA: MIT Press.

Wheeler, J. A. (1990). Information, physics, quantum: The search for links. In W. H. Zurek (Ed.), *Complexity, entropy, and the physics of information*. Redwood City: Addison-Wesley.

Wheeler, J., & Zurek, W. (1983). *Quantum theory and measurement*. Princeton: Princeton Univ. Press.

Whitehead, A. N. (1927). *Symbolism: Its meaning and effect*. New York: Macmillan.

Wigner, E. P. (1964). Events, laws, and invariance principles. *Science, 145*, 995–999.

Yates, F. E. (1987). *Self-organizing systems: The emergence of order*. New York: Plenum.

Yovits, M. C., & Cameron, S. (Eds.). (1960). *Self organizing systems*. New York: Pergamon.

Yovits, M. C., Jacobi, G. T., & Goldstein, G. D. (Eds.). (1962). *Self organizing systems, 1962*. Washington, DC: Spartan.

Zurek, W. H. (Ed.). (1990). *Complexity, entropy, and the physics of information*. Redwood City: Addison-Wesley.

14 Artificial Life Needs a Real Epistemology

Life is peculiar, said Jeremy. As compared with what? asked the spider.

(Curtis and Greenslet 1945)

Abstract Foundational controversies in artificial life and artificial intelligence arise from lack of decidable criteria for defining the epistemic cuts that separate knowledge of reality from reality itself, e.g., description from construction, simulation from realization, mind from brain. Selective evolution began with a description-construction cut, i.e., the genetically coded synthesis of proteins. The highly evolved cognitive epistemology of physics requires an epistemic cut between reversible dynamic laws and the irreversible process of measuring initial conditions. This is also known as the measurement problem. Good physics can be done without addressing this epistemic problem, but not good biology and artificial life, because open-ended evolution requires the physical implementation of genetic descriptions. The course of evolution depends on the speed and reliability of this implementation, or how efficiently the real or artificial physical dynamics can be harnessed by non-dynamic genetic symbols.

14.1 What Can Artificial Life Tell Us About Reality?

When a problem persists, unresolved, for centuries in spite of enormous increases in our knowledge, it is a good bet that the problem entails the nature of knowledge itself. The nature of life is one of these problems. Life depends on matter, but life is not an inherent property of matter. Life is peculiar, obviously, because it is so different from nonliving matter. It is different, not so obviously, because it realizes an intrinsic

---

Reprinted from *Advances in Artificial Life*, F. Moran, A. Moreno, J. J. Merelo, O. Chacon, Eds. Berlin: Springer, 1995, pp. 23–38.

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7, DOI 10.1007/978-94-007-5161-3_15, © Springer Science+Business Media Dordrecht 2012

epistemic cut between the genotype and phenotype. Our knowledge of physics, chemistry, molecular biology, genetics, development, and evolution is enormous, but the question persists: Do we really understand how meaning arises from matter? Is it clear why nonliving matter following inexorable universal laws should acquire symbolic genes that construct, control, and evolve new functions and meanings without apparent limit? In spite of all this knowledge, most of us still agree with Jeremy.

Where we find disagreement is on the answer to the spider's question. Artificial life must ask this question: With what do we compare artificial life? The founding characterizations of artificial life comparing "*life-as-it-could-be*" with "*life-as-we-know-it*" (Langton 1989; Langton 1992), or "implementation-independent" computer life with space-time-energy-dependent material life (Fontana et al. 1994) was a creative beginning, but this highly formal view of life was immediately questioned. Such abstract characterizations do not clearly separate science fiction and computer games from physical reality (Pattee 1988). On the other hand, the idea of life in a computer does stimulate the philosophical imagination.

An alternative view of artificial life uses computation to control robots in a real physical world. Although in this approach the more fundamental philosophical issues are not as apparent, it has the enormous advantage in a practical sense of using the physical world at face value. As Brooks (1992) understates the point: "It is very hard to simulate the actual dynamics of the real world."

My first answer to the spider's question is that we can only compare life to non-life, that is, to the nonliving world from which life arises and evolves. Artificial life must be compared with a real or an artificial nonliving world. Life in an artificial world requires exploring what we mean by an alternative physical or mathematical reality. I want to follow Dennett's (1994) suggestion that we use artificial life as a "prosthetically controlled thought experiment" that may provide some insights into such foundational questions. Metaphysical questions, like whether reality is material, formal, or mental, are empirically undecidable, but nevertheless, discussion of these concepts are an important part of scientific discovery. Historically we have seen concepts of reality shifting ground and new horizons discovered, especially with the advent of quantum theory, computation theory, and cosmology. The question is: Can artificial life add some new ideas to the problem of knowledge and the epistemic cut or will it only increase the confusion?

14.2 Life Requires an Epistemic Cut

The first problem for life in a computer is to recognize it. How peculiar does artificial life have to be? That is, how will we distinguish the living parts of the computation from the nonliving parts? And what are the parts? I have argued for many years that life is peculiar, fundamentally, because it separates itself from nonliving matter by incorporating, within itself, autonomous epistemic cuts (Pattee 1969, 1972, 1982, 1993, 1995). Metaphorically, life is matter with meaning. Less metaphorically, organisms are material structures with memory by virtue of which they construct, control and adapt to their environment. Evolution entails semantic information (Eigen 1992), and open-ended evolution requires an epistemic cut between the genotype and phenotype, i.e., between description and construction. The logical necessity of this epistemic cut is the fundamental point of von Neumann's (1966) self-replicating automaton. It is this type of *logical* argument that gives some validity to the concept of formal life, implementation-independent life, or life in a computer.

It is not clear how far such logical arguments can take us. As von Neumann warned, if one studies only formal life, "(…) one has thrown half the problem out the window and it may be the more important half." In spite of all our knowledge of the chemical properties of the components of the genotype and phenotype, no one knows the answer to von Neumann's "most intriguing, exciting and important question of why the molecules or aggregates which in nature really occur (…) are the sorts of things they are." In fact, this question is the best reason I know for studying artificial life where we can invent different "sorts of things" and see how they behave.

14.2.1 The Epistemic Cut Requires Implementation

What does implementing a description mean? Descriptions are nondynamic stored structures that do nothing until they are interpreted and implemented. In life-as-we-know-it this means *translating and constructing what is described*. We know that this is a very complex process in real life involving DNA, messenger RNA, transfer RNA, coding enzymes, ribosomes, and a metabolism to drive the entire synthesis process. It is therefore not clear what total implementation-independence or formalization of artificial life can tell us. It is precisely the effectiveness of implementation of genetic descriptions that evolution by natural selection is all about. Complete formalization would indeed throw half the problem out the window, as von Neumann says.

The central problem of artificial life, as theoretical biology, is to separate the essential aspects of this implementation from the frozen accidents or the incidental chemistry and physics of the natural world that might have been otherwise. Of course all these levels of detail are useful for the problems they address, but to answer the question of why these molecules are the "sorts of things they are" requires abstracting just the right amount.

It is not generally appreciated in artificial life studies why formal self-replication is only half the problem. All of evolution, emergence, adaptation, and extinction, depends on how quickly and efficiently the variations in the genotype can be implemented in phenotypic functions. How does a symbolic sequence-space map into a physical function-space? In spite of all the physical and chemical knowledge we have, it still appears unreasonably fortuitous that only linear sequences of nucleotides are sufficient to instruct the synthesis of all the structural proteins including their self-folding and self-assembling properties, and all the coordinated, highly specific and powerful enzymes that control the dynamics of all forms of life. It is significant that even at the simplest level the implementation entails a computationally intractable problem—the polymer folding problem.

The advantage of the autonomous robotics approach to artificial life is that it avoids the most intractable computational problems in the same way that real life does—it harnesses the real physics. However, robotics does not face the more fundamental construction and self-assembly problems. The question is how much can we learn from computational models alone about such efficient implementations of genetic information? Such questions depend largely on our epistemology of computation, that is, how we think of measurements and symbols constraining a dynamics. The same problem exists for all physical systems, as I will discuss in Sect. 14.3. I will survey some current concepts of computation after outlining what I mean by an epistemology and summarizing the standard epistemic principles of physical theory.

In traditional philosophy epistemic cuts are viewed as problems only at the cognitive level. They are called problems of reference or how symbols come to "stand for" or to "be about" material structures and events (Cassirer 1957; Harnad 1990; Whitehead 1927). I have always found the complementary problem much more fundamental: How do material structures ever come to be symbolic? I think if we fully understood how molecules become messages in cells we would have some understanding of how messages have meaning. That is why the origin of life problem is important for philosophy.

14.3 What Is an Epistemology?

An epistemology is a theory or practice that establishes the conditions that make knowledge possible. There are many epistemologies. Religious mystics, and even some physicists (Wilber 1985), believe that higher knowledge is achieved by a state of ineffable oneness with a transcendent reality. Mystics do not make epistemic cuts. While this may work for the individual, it does not work for populations that require *heritable* information or common knowledge that must be communicable (Born 1964). Knowledge is potentially useful information *about* something. Information is commonly represented by *symbols*. Symbols *stand for* or are *about* what is represented. Knowledge may be about what we call reality, or it may be about other knowledge. It is the *implementation* of "standing for" and "about"—the process of executing the epistemic cut—that artificial life needs to explore.

Heritable, communicable, or objective knowledge requires an epistemic cut to distinguish the knowledge from what the knowledge is about. By *useful* information or knowledge I mean information in the evolutionary sense of information for construction and control, measured or selected information, or information ultimately necessary for survival. This is contrasted with ungrounded, unmeasured, unselected, hence, purely formal or syntactic information. My usage does not necessarily imply higher-level cognitive concepts like understanding and explanation, neither does it exclude them. I am not troubled by the apparent paradox that primitive concepts may be useful without being precisely understood. I agree with C. F. von Weizsäcker (1973), "Thus we will have to understand that it is the very nature of basic concepts to be practically useful without, or at least before, being analytically clarified."

14.3.1 The Epistemology of Physical Theory

The requirement for heritable or objective knowledge is the separation of the subject from the object, the description from the construction, the knower from the known. Hereditary information originated with life with the separation of description and construction, and after 3.6 billion years of evolution this separation has developed into a highly specialized and explicit form at the cognitive level. Von Neumann (1955) states this epistemology of physical theory clearly: "(…) we *must* always divide the world into two parts, the one being the observed system, the other the observer. The boundary between the two is arbitrary to a very large extent (…) but this does not change the fact that the boundary *must* be put somewhere, if the method is not to proceed vacuously (…)" In physical theory, the observer is *formally* related to the observed system only by the *results* of measurements of the observables defined by the theory, but the formulation of the theory, the choice of observables, the construction of measuring devices, and the measurement process itself cannot be formalized.

No matter where we divide the world into observed and observer, the fundamental condition for physical laws is that they are invariant to different observers or to the frames of reference or states of observers. Laws therefore hold everywhere—they are universal and inexorable. In addition to the invariance or symmetry principles, the laws must be separated from the initial conditions that are determined only by measurement. The distinction between laws and initial conditions can also be expressed in terms of information and algorithmic complexity theory (Chaitin 1987). Algorithmic complexity of information is measured by the shortest program on some Turing-like machine that can compute this information. Laws then represent information about the world that can be enormously shortened by algorithmic compression. Initial conditions represent information that cannot be so compressed.

Mystical and heritable epistemologies are not necessarily incompatible. They simply refer to different forms of knowledge (Eddington 1928). For example, Penrose (1989) agrees that this separation of laws is "historically of vital importance" but then expresses more mystically the "very personal view" that "when we come ultimately to comprehend the laws…this distinction between laws and boundary conditions will dissolve away."

14.3.2 Incomplete Knowledge—The Necessity of Statistical Laws

The epistemology of physics would be relatively simple if this were all there were to it, but laws and initial conditions alone are not enough to make a complete physical theory that must include measurement. Measurement and control require a third category of knowledge called boundary conditions or constraints. These are initial conditions that can be compressed *locally* but that are neither invariant nor universal like laws. When such a constraint is viewed abstractly it is often called a rule; when it is viewed concretely it is often called a machine or hardware.

Both experience and logic teach us that initial conditions cannot be *measured*, nor boundary conditions *constructed*, with the deterministic precision of the formal dynamical laws. Consequently, this third category of knowledge requires *statistical laws*. Statistical laws introduce one of the great unresolved fundamental problems of epistemology. The dynamical laws of physics are all symmetric in time and therefore reversible, while statistical laws are irreversible. Formally, these two types of laws are incompatible. It is even difficult to relate them conceptually. From Bernoulli and Laplace to the present day this problem persists. As Planck (1960) says, “For it is clear to everybody that there must be an unfathomable gulf between a probability, however small, and an absolute impossibility.” He adds, “Thus dynamics and statistics cannot be regarded as interrelated.” Von Neumann (1966) agreed with Planck but cautioned, “(…) the last word about this subject has certainly not been said and it is not going to be said for a long time.” Thirty years later, Jaynes (1990) says about the interpretation of probability in quantum theory, “(…) we are venturing into a smoky area of science where nobody knows what the real truth is.”

What types of boundary conditions or constraints can “self-organize” from deterministic dynamical laws, and what types can only “emerge” from a statistical bias on a heritable population distribution (i.e., natural selection) is a central problem in evolution theory and an active study in artificial life (Langton 2999). As with all such problems, the issue depends on the existence of an epistemic cut.

14.3.3 Measurement Defines an Epistemic Cut

Like it or not, the epistemic cut in physical theory falls in Planck’s “unfathomable gulf” between dynamical and statistical laws. The possible trajectories of the world are described dynamically by reversible, noiseless laws, but any explicit knowledge of a trajectory requires observations or measurements described by irreversible, noisy statistical laws. This is the root of the measurement problem in physical theory. The problem arises classically, where it is often discussed using the thought experiments such as Maxwell’s demon (Leff and Rex 1990), and in quantum theory where the formal treatment of the measurement process only makes matters worse (Wheeler and Zurek 1983). Von Neumann (1955) described the problem in this way: An epistemic cut must separate the measuring device from what is measured. Nevertheless, the constraints of the measuring device are also part of the world. The device must therefore be describable by universal dynamical laws, but this is possible only at the cost of moving the epistemic cut to exclude the measurement. We then require a new observer and new measuring devices—a vacuous regress.

When we distinguish the Turing-von Neumann concept of programmable computation from other less well-defined concepts, we will see in Sect. 14.5.6 that when described physically a “step” in the computation must be a measurement. The completion of a measurement is indicated by a record or memory that is no longer a part of the dynamics except as an incoherent (nonintegrable) constraint.

It is important to understand that invariance and compressibility are not themselves laws, but are necessary epistemic conditions to establish the heritability, objectivity and utility we require of laws. As P. Curie (1908) pointed out, if the entire world in all its details were really invariant there would be nothing to observe. No epistemic cut would be possible, and therefore life could not exist, except perhaps in a mystical sense. It is only because we divide our knowledge into two categories, dynamical laws and initial conditions, that invariance itself has any meaning (Houtappel et al. 1965; Wigner 1964). How we choose this cut intellectually is largely a pragmatic empirical question, although there is also a strong aesthetic component of choice (Polanyi 1964).

The point is that invariance and compressibility are general epistemic requirements for evolution that preceded physical theory. They are both “about” something else, and therefore they require a cut between what does not change and what does change, and between the compression and what is compressed. How life, real or artificial, spontaneously discovers an invariant, compressible, and hence evolvable, description-construction cut is the origin of life problem. However it happened, it is clear that compressibility is necessary to define dynamical laws and life. Without compressibility life could not adapt or evolve, because there is no way to adapt to endlessly random (incompressible) events.

14.4 Artificial Life Requires an Artificial Physics

How is this physical epistemology relevant for artificial life? The important point is that physical epistemology is a highly evolved and specialized form of the primitive description-construction process. The cognitive role of physical epistemology appears to be far removed from the constructive function of genes, but both define a fundamental epistemic cut. Great discoveries have been made in physics without understanding the mechanisms that actually implement the epistemic cut, because physics does not need to study the epistemic cut itself. Measurement can simply be treated as an irreducible primitive activity. That is why in most sciences the epistemic cut appears sharp—we tend to ignore the details of constructing the measurement devices and record only the results. The reality is that physical theory would remain in a primitive state without complex measuring devices, and in fact most of the financial resources in physics are spent on their construction.

Unlike physical theory, great discoveries in the evolution of natural and artificial life are closely related to understanding how the description-construction process can be most efficiently *implemented*. The course of evolution depends on how rapidly and efficiently an adaptive genotype-phenotype transformation can be discovered and how reliably it can be executed (Conrad 1983, 1989).

Real and artificial life must have arisen and evolved in a nonliving milieu. In real life we call this the real physical world. If artificial life exists in a computer, the computer milieu must define an artificial physics. This must be done explicitly or it will occur by default to the program and hardware. What is an artificial physics or physics-as-it-might-be? Without principled restrictions this question will not inform philosophy or physics, and will only lead to disputes over nothing more than matters of taste in computational architectures and science fiction. If an epistemology-as-we-know-it in physics has evolved from life itself, we must consider this a fundamental restriction. What we now distinguish as the three essential categories of knowledge—laws, initial conditions, and statistics—we need to represent in computational models of artificial life.

This means that artificial laws must correspond to algorithmically compressible subsets of computational events, and initial conditions must refer to incompressible events determinable only by measurements by organisms. In other words, any form of artificial life must be able to detect events and discover laws of its artificial world. Defining a measurement in a computer is a problem. I discuss this in Sect. 14.5.6. Also, autonomy requires what I call *semantic closure* (Pattee 1995). This means the organism's measurement, memory, and control constraints must be *constructed* by the genes of the organism from parts of the artificial physical world. Of course consistency requires that all activities of the organisms follow the laws they may discover. Whether such organisms are really alive or only simulated is a matter of definition. A more objective and important question is how open-ended is such computational life. No consensus can be expected on this question unless there is consensus on what computation means.

14.5 What Is Computation?

There are two fundamentally different views of computation, the mathematical or formal view and the physical or hardware view. Barrow (1991) sees these views arising from "two great streams of thought" about physical reality. The traditional view is based on symmetry principles, or invariance with respect to observers. The currently popular view is based on an abstract concept of computation. Roughly, the symmetry view is based on the established physical epistemology that I outlined above with statistical measurement playing an essential role. The computational view emphasizes a dynamical ontology, with logical consistency and axiomatic laws playing the essential role. The one view sees computation as a locally programmable, concrete, material process strictly limited by the laws of physics. The other view sees computation as a universal, abstract dynamics to which even the laws of physics must conform.

14.5.1 Formal Computation

The ontological view of computation has some roots in the historical ideal of formal symbol manipulation by axiomatic rules. The meaning of a formal system in logic and mathematics as conceived by Hilbert is that all the procedures for manipulating symbols to prove theorems and compute functions are axiomatically specified. This means that all the procedures are defined by idealized unambiguous rules that do not depend on physical laws, space, time, matter, energy, dissipation, the observer's frame of reference, or the many possible semantic interpretations of the symbols and rules. The founders of computation theory were mostly logicians and mathematicians who, with the significant exceptions of Turing and von Neumann, were not concerned with physical laws. Ironically, formal computational procedures are now called "effective" or "mechanical" even though they have no epistemic relation to physical laws. These procedures are justified only by intuitive thought experiments. This weakness is well-known, but is usually ignored. As Turing (1936) noted, "All arguments which can be given [for effective procedures] are bound to be, fundamentally, appeals to intuition and for this reason rather unsatisfactory mathematically."

This complete conceptual separation of formal symbol manipulation from its physical embodiment is a characteristic of mathematical operations as we normally do them. All symbol strings are discrete and finite, as are all rewriting steps. Steps may not be analyzed as analog physical devices. Proofs do not allow statistical fluctuation and noise. The concepts of set and function imply precise symbol recognition and complete determinism in rewriting all symbols. Formal computation is, *by definition*, totally implementation-independent (Kleene 1952).

This formal view of computation appears to contribute little to understanding the nature of epistemic cuts because formal systems are self-contained. Symbols and rules have no relation to measurement, control, and useful information. In fact, purely formal systems must be free of all influence other than their internal syntax, otherwise they are in error. To have meaning they must be informally interpreted, measured, grounded, or selected *from the outside*. "Outside" of course is established only by an epistemic cut. It is for this reason that formal models can be programmed to *simulate* everything, except perhaps the ineffable or mystical, since all the interpretation you need to define what the simulation means can be freely provided from outside the formal activity of the computer.

14.5.2 Laplacean Computation

An extension of formal computation is the Laplacean ideal which, as a thought experiment, replaces the epistemic cut with an in-principle isomorphism between the formal computational states and physical states. Such thought experiments often lead to apparent paradoxes precisely because an isomorphism is a formal concept that does not define how to execute the epistemic cut. Maxwell's demon and Schrödinger's cat are famous examples. The demon forces us to clearly state how measured information is distinguished from physical entropy, and the cat forces us to decide when a measurement occurs. These distinctions both require defining epistemic cuts between the knower and the known. It is significant that there is still no consensus on where the cut should be placed in both cases (Leff and Rex 1990; Wheeler and Zurek 1983).

14.5.3 Computation in the Wild

A further elaboration of formal computation has become popular more recently as a kind of backward Laplacean ideal. That is, the Laplacean isomorphism is interpreted as its converse: computation does not provide a map of the universe—the universe is a map of a computation, or "IT from BIT" as Wheeler (1990) states it. This ontological view is what Dietrich (1994) calls "the computer in the wild." Historians might try to blame this view on Pythagoras, but its modern form probably began with the shift in the view of mathematics as a pure logical structure to more of a natural science after the failure of pure logic to justify its foundations. The ontological view also arose from the ambiguous relation of information to entropy in the contexts of cosmology, quantum theory, and algorithmic complexity theory (Chaitin 1987; Zurek 1990). Toffoli (1982) describes computation this way: "In a sense, nature has been continually computing the 'next state' of the universe for billions of years; all we have to do—and actually all we can do—is 'hitch a ride' on this huge ongoing computation, and try to discover which parts of it happen to go near where we want."

This confounding of formal rules that arise from constraints, and dynamics that describe physical laws, leads to ambiguous questions like, "Is there a physical phenomenon that computes something noncomputable? Contrariwise, does Turing's thesis…constrain the physical universe we are in?" (Chaitin, 1982). This speculative association of formal theorems with physical laws is sometimes called the strong Church-Turing thesis. It leads to the argument that if there were a natural physical process that could not be Turing-computed, then that process could be used as a new computing element that violates the thesis (Conrad 1989).

The strong Turing-Church thesis is also used to try to equate formal Turing-equivalence between two symbol systems, with fitness equivalence between two physical implementations of the formal systems. The argument is that because there are many Turing-equivalent formalisms, like cellular automata and artificial neural nets, that there is no significant difference in the behavior of their different physical implementations. Of course from the biological perspective this is not the case, because it is precisely the overall *efficiency* of the physical implementation that determines survival. The significant processes in life at all levels, from enzyme catalysis to natural selection, depend on statistical biases on the rates of change of noisy population distributions, whereas formal equivalence is neither a statistical bias, rate-dependent, noisy, nor a population distribution.

The believers in strong artificial intelligence have further popularized the computer metaphor by defining brains and life as just some kind of computer that we do not yet understand. This view is labeled *computationalism*. It replaces the Laplacean isomorphism with an identity. Like Toffoli, Dietrich (1994) believes that, "every physical system in the universe, from wheeling galaxies to bumping proteins, is a special purpose computer in the sense that every physical system in the universe is implementing some computation or other." According to this view, the brain is a computer *by definition*. It is our job to figure out what these physical systems are really computing. Thus, according to Churchland and Sejnowski (1992), "…there is a lot we do not yet know about computation. Notice in particular that once we understand more about what sort of computers *nervous systems* are, and how they do whatever it is they do, we shall have an enlarged and deeper understanding of what it is to compute and represent." Hopfield (1994) extends this vague, generalized view of computation to evolution: "Much of the history of evolution can be read as the evolution of systems to make environmental measurements, make predictions, and generate appropriate actions. This pattern has the essential aspects of a computational system."

This undifferentiated view of the universe, life, and brains as all computation is of no value for exploring what we mean by the epistemic cut because it simply includes, *by definition*, and without distinction, dynamic and statistical laws, description and construction, measurement and control, living and nonliving, and matter and mind as some unknown kinds of computation, and consequently misses the foundational issues of what goes on within the epistemic cuts in all these cases. All such arguments that fail to recognize the necessity of an epistemic cut are inherently mystical or metaphysical and therefore undecidable by any empirical or objective criteria (Pattee 1989, 1990; Rosen 1986).

14.5.4 The Programmable Physical Computer

The formal view of computation would be conceivable as long as Turing's (1936) condition that every symbol is "immediately recognizable" (i.e., perfectly precise measurement) and Gödel's (1964) condition of perfect mechanism (i.e., perfect determinism) were possible. However, even though we have no way of knowing if nature is ultimately deterministic or not, we do know that measurement must at some stage be irreversible, and the results of measurement cannot be used to violate the 2nd law of thermodynamics. Hence, useful measurements are dissipative and subject to error and violate the assumptions of Laplace, Turing, and Gödel.

The physical view of computation is little more than the engineering view that recognizes the hardware constraints as a necessity for implementing any symbol manipulation. This view holds that statistical physical laws are both the foundation and limitation of computation. Programmable hardware is inherently slow and noisy. Most of the peculiar design features of the computer are to overcome these limits. It is a wonder of technology that these limits have been extended so far. Actually, it was Turing (1936) who first justified the use of bits as the highest signal-to-noise symbol vehicle, and of course von Neumann (1966) believed that any rigorous theory of computation must have its roots in thermodynamics. He did not think of computers as implementation-independent: "An automaton cannot be separated from the milieu to which it responds. By that I mean that it is meaningless to say that an automaton is good or bad, fast or slow, reliable or unreliable, without telling in what milieu it operates." The same is true for natural and artificial life.

14.5.5 Limits of Physical Computation

The requirement that computation must satisfy physical laws, especially the 2nd law of thermodynamics, is seldom questioned, but is nevertheless largely ignored by both formalists and computationalists. On the other hand, hardware designers are acutely aware of the practical physical limits of speed, reliability and dissipation. Theories of reversible (dissipationless) computation have been proposed over the last few decades (Bennett 1982; Landauer 1986), but they are essentially thought experiments with idealized dynamical constraints. No one knows how to build a useful programmable computer along these lines.

Bennett, (1987) argues that the source of irreversibility in measurement is erasure rather than the measurement itself. This interpretation is possible if the measured results remain unused on the physical side of the epistemic cut. In any case, the basic condition is that our use of measured information cannot lead to a violation of the second law of thermodynamics. Therefore the addition of any new measured information that is actually used to decrease the accessible states (i.e., entropy decrease, useful work, natural selection, control of dynamics, etc.) must be compensated by information loss (i.e., entropy increase, noise, dissipation, increased accessible states, etc.) in some aspect of the measuring process (Zurek 1990).

14.5.6 Analog Dynamics

The ontological computationalists will argue that normal programmable computation is just our *interpretation* of a constrained physical dynamical system. They claim that all dynamical systems, likewise, can be interpreted as computing because they are *implementing* some functions (Dietrich 1994; Toffoli 1982). This may be the case for analog computers where the dynamics maps initial states to final states without programs, measurements, or intermediate steps, but this is too great an abstraction for describing a programmable computer.

The issue is what we mean by implementation of a function, and how we define a step. If a computer is to be an implementation of formal logic or mathematics, then it must implement *discrete symbols* and perform *discrete steps* in the rewriting of these symbols to and from memory according to a sequence of rules or a program. This is what formal logic and mathematics is about. This is what Turing/von Neumann programmable computers do. It is also the case that any implementation of such symbolic computational steps must be a law-based system with physical dynamics, so the question is: What corresponds to a symbol and a step? Physical dynamics does not describe symbols and steps. They are not in the primary categories of knowledge called laws and initial conditions. A *step* can only be defined by a measurement process, and a *symbol* as a record of a measurement. Therefore, a programmable computation can be described in physical terms only as a dynamical system that is internally constrained to regularly perform a sequence of simple measurements that are recorded in memory. The records subsequently act as further constraints. Since the time of measurement, by definition, has no coherence with the time of the dynamics, the sequence of computational steps is *rate-independent*, even though all physical laws are *rate-dependent*. As in all arguments about when measurement occurs, this also depends on where the epistemic cut is placed.

The ontological computer-in-the-wild is a physical system that may be interpreted as a dynamical analog device that parallels some other process. Such analog computers were common before the development of the programmed digital computer. They cannot be classified easily because all systems are indeed potential analogs. Furthermore, what aspects of the system are to be interpreted as computation are not crisply defined as are symbolic, rule-based, programmed computers. It should be clear that these are two extremes that only produce confusion by being lumped together. In rate-dependent dynamical analogs no memory is necessary, and one epistemic cut is made at the end when the final result is measured. In rate-independent programmed computation each step is a measurement recorded in memory. There are innumerable possibilities for machines with all degrees of constraints in between these extremes, but none have general utility.

14.6 The Epistemology of Organisms

Living systems as-we-know-them use a hybrid of both discrete symbolic and physical dynamic behavior to implement the genotype-phenotype epistemic cut. There is good reason for this. The source and function of genetic information in organisms is different from the source and function of information in physics. In physics new information is obtained only by measurement and, as a pure science, used only passively, to know *that* rather than to know *how*, in Ryle's terms. Measuring devices are designed and constructed based on theory. In contrast, organisms obtain new genetic information only by natural selection and make active use of information to know *how*, that is, to construct and control. Life is constructed, but only by trial and error, or mutation and selection, not by theory and design. Genetic information is therefore very expensive in terms of the many deaths and extinctions necessary to find new, more successful descriptions.

This high cost of genetic information suggests an obvious principle that there is no more genetic information than is necessary for survival. What affects this minimum? The minimum amount of genetic or selected information will depend largely on how effectively this information can be implemented using the parts and the dynamics of physical world. For example, some organisms require genetic instruction for synthesizing amino acids from smaller molecules, but if all amino acids are available as environmental parts, there is no need for these genes. At the next level, if the information that determines the linear sequence is sufficient constraint to determine the folding and self-assembly of proteins then no further folding information is necessary. However, in some cases, when the self-folding is not stable, additional genes for membrane or scaffolding proteins to further constrain the folding are necessary.

This minimum genetic information principle should not be confused with algorithmic compression of information. Algorithmic compression is defined only in a formal context on unselected information. Compressibility across an epistemic cut can only be interpreted informally as the efficiency of implementation of selected information in a physical milieu. No such minimum information principle can apply to formal or programmable computation. Formal computation requires, by definition, complete informational control. That is, no self-folding or any other law-based dynamics can have any effect on formal symbol manipulation. Any such effect is regarded either as irrelevant or an error.

The success of evolution depends on how quickly and effectively organisms can adapt to their environment. This in turn depends on how efficiently the sequence space of genes can transform, gradually, the control or function space of phenotypes. Efficiency here includes the search problem, i.e., how to find good descriptions (Schuster 1994), and the construction problem, i.e., how to reliably assemble parts according to the description (Conrad 1983, 1990).

As I mentioned in Sect. 14.2.1, it is important to recognize that these implementation problems, if treated formally, are combinatorially complex. The search space is enormous and the number of degrees of freedom of an enzyme is large, so that even though polymer folding is the simplest possible natural process that implements the genotype-phenotype transformation, a purely computational mapping is impractical. Even the two-dimensional folding of RNA is NP-complete, and *ab initio* computation of detailed protein folding appears out of reach. To make matters worse, folding requires finding only a stationary state. A quantum dynamical model of enzyme catalysis has not even been formulated.

The only practical computational approach to these combinatorially complex problems is to use “reverse biological engineering” and simulate the natural dynamics with artificial neural nets (Hunter 1993; Zuker 1989), and natural selection in the form of genetic algorithms to evolve the connection weights in the nets (Whitley and Hanson 1989). There is no doubt that these techniques derived from life-as-we-know-it are of practical engineering value. However, I would call them virtual dynamical analogs implemented by programmed computers. Adlelman (1994) has used real DNA molecules in a “massively parallel” chemical search for a solution of the Hamiltonian path problem. It is a matter of taste whether this should be called molecular computing or chemical graph theory.

14.7 Conclusions

If artificial life is to inform philosophy, physics, and biology it must address the implementation of epistemic cuts. Von Neumann recognized the *logical* necessity of the description-construction cut for open-ended evolvability, but he also knew that a completely axiomatic, formal, or implementation-independent model of life is inadequate, because the course of evolution depends on the *speed*, *efficiency*, and *reliability* of implementing descriptions as constraints in a dynamical milieu.

Many nonlinear dynamical models of populations of interacting units, like cellular automata, Kauffman-type networks, and games of life have been interpreted as genetic populations with or without a genotype-phenotype mapping. These populations compete, cooperate, and coevolve in an artificial environment. However, where there is a genotype-phenotype mapping this is usually a fixed program in which the evolution of efficiency of implementation does not arise. Of course implementation-independent self-organization may play essential roles in the origin of life and in limiting the possibilities for natural selection. The significance of these roles needs to be determined.

Implementation of a description means constructing organisms with the parts and the laws of an artificial physical world. Some epistemic principles must restrict physics-as-it-could-be if it is to be any more than computer games. In order to evolve, organisms must discover by selection or measurement some "compressible" genetic descriptions of this artificial physical world. Compressibility is a formal concept that is not strictly applicable across epistemic cuts. Compressibility across an epistemic cut simply corresponds to efficiency of implementation. There is no general way to measure how far we can compress the genetic information that is necessary to implement a biological function, because this depends on the physical laws and the quality of function necessary for survival. For the same reason we cannot in general specify how much information is necessary to construct a measuring device. The amount of information for construction of the measuring device is incommensurable with the survival value of the information obtained by the measurement. This is generally the case for all biological structure-function relations.

To evolve, organisms must efficiently implement these descriptions as constructions. A fundamental limitation for computer life is that evolution can only reflect the complexity of the artificial physical world in which organisms live. An epistemic cut affords the *potential* for efficient implementation and open-ended evolution, but in a simple world, efficient implementations will be limited and life will also remain simple.

Hybrid symbolic-dynamic systems, like life-as-we-know-it and computer-controlled robots actually address the problem of efficient implementation of control instructions in the real world, but robots are still a long way from implementing efficient memory-controlled *constructions* of real life that self-assemble at all levels, from polymer folding to multicellular development. Real world dynamics will always have some advantages for efficient implementations, because there are necessary but gratuitous inefficiencies of programmed computer simulations that are missing in reality, as well as significant unpredictable efficiencies of reality missing in the simulations.

References

Adelman, L. M. (1994). Molecular computation of solutions to combinatorial problems. *Science, 266*, 1021–1024.

Barrow, J. D. (1991). *Theories of everything* (p. 203). New York: Oxford University Press.

Bennett, C. H. (1982). The thermodynamics of computation—A review. *International Journal of Theoretical Physics, 21*, 905–940.

Bennett, C. H. (1987). Demons, engines, and the second law. *Scientific American, 257*, 108–116.

Born, M. (1964). Symbol and reality. *Universitas 7*, 337–353. Reprinted in Born, M. *Physics in my generation* (pp. 132–146). New York: Springer.

Brooks, R. (1992). Artificial life and real robots. In F. Varela & P. Bourgine (Eds.), *Toward a practice of autonomous systems* (pp. 3–10). Cambridge, MA: MIT Press.

Cassirer, E. (1957). *The philosophy of symbolic forms* (The phenomena of knowledge, Vol. 3). New Haven: Yale University Press.

Chaitin, G. (1982). Gödel's theorem and information. *International Journal of Theoretical Physics, 21*, 941–953.

Chaitin, G. (1987). *Algorithmic information theory*. Cambridge/ New York: Cambridge University Press.

Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain*. Cambridge, MA: MIT Press.

Conrad, M. (1983). *Adaptability*. New York: Plenum.

Conrad, M. (1989). The brain-machine disanalogy. *BioSystems, 22*, 197–213.

Conrad, M. (1990). The geometry of evolution. *BioSystems, 24*, 61–81.

Curie, P. (1908). *Oeuvres* (p. 127). Paris: Gauthier-Villars.

Curtis, C., & Greenslet, F. (1945). *The practical cogitator* (p. 277). Boston: Houghton Mifflin.

Dennett, D. (1994). Artificial life as philosophy. *Artificial Life, 1*, 291–292.

Dietrich, E. (1994). *Thinking computers and virtual persons* (p. 13). New York: Academic Press.

Eddington, A. (1928). *The nature of the physical world* (p. 260). New York: Macmillan.

Eigen, M. (1992). *Steps toward life*. Oxford: Oxford University Press.

Fontana, W., Wagner, G., & Buss, L. W. (1994). Beyond digital naturalism. *Artificial Life, 1*, 211–227.

Gödel, K. (1964). Russell's mathematical logic, and what is Cantor's continuum problem? In P. Benacerraf & H. Putnam (Eds.), *Philosophy of mathematics* (pp. 211–232). Englewood Cliffs: Prentice-Hall, 258–273.

Harnad, S. (1990). The symbol grounding problem. *Physica D, 42*, 335–346.

Hopfield, J. J. (1994). Physics, computation, and why biology looks so different. *Journal of Theoretical Biology, 171*, 53–60.

Houtappel, R. M. F., Van Dam, H., & Wigner, E. P. (1965). The conceptual basis and use of the geometric invariance principles. *Reviews of Modern Physics, 37*, 595–632.

Hunter, L. (Ed.). (1993). *Artificial intelligence and molecular biology*. Menlo Park: AAAI/MIT Press.

Jaynes, E. (1990). Probability in quantum theory. In W. H. Zurek (Ed.), *Complexity, entropy, and the physics of information* (p. 382). Redwood City: Addison-Wesley.

Kleene, S. C. (1952). *Introduction to metamathematics*. Princeton: Van Nostrand.

Landauer, R. (1986). Computation and physics: Wheeler's meaning circuit. *Foundations of Physics, 16*, 551–564.

Langton, C. (1989). Artificial life. In C. Langton (Ed.), *Artificial life* (pp. 1–47). Red-wood City: Addison-Wesley.

Langton, C., Taylor, C., Farmer, J., & Rasmussen, S. (Eds.). (1992). *Artificial life II*. Redwood City: Addison-Wesley.

Leff, H. S., & Rex, A. F. (Eds.). (1990). *Maxwell's demon, entropy, information, computing*. Princeton: Princeton University Press.

Pattee, H. H. (1969). How does a molecule become a message? *Developmental Biology Supplement, 3*, 1–16.

Pattee, H. H. (1972). Laws, constraints, symbols, and languages. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Vol. 4, pp. 248–258). Edinburgh: Edinburgh University Press.

Pattee, H. H. (1982). Cell psychology: An evolutionary approach to the symbol-matter problem. *Cognition and Brain Theory, 5*(4), 325–341.

Pattee, H. H. (1988). Simulations, realizations, and theories of life. In C. Langton (Ed.), *Artificial life* (pp. 63–77). Redwood City: Addison-Wesley.

Pattee, H. H. (1989). The measurement problem in artificial world models. *BioSystems, 23*, 281–290.

Pattee, H. H. (1990). Response to E. Dietrich's "Computationalism". *Social Epistemology, 4*(2), 176–181.

Pattee, H. H. (1993). The limitations of formal models of measurement, control, and cognition. *Applied Mathematicals and Computation, 56*, 111–130.

Pattee, H. H. (1995). Evolving self-reference: Matter, symbols, and semantic closure. *Communication and Cognition - AI, 12*(1–2), 9–28.

Penrose, R. (1989). *The emperor's new mind* (p. 352). Oxford: Oxford University Press.

Planck, M. (1960). *A survey of physical theory* (p. 64). New York: Dover.

Polanyi, M. (1964). *Personal knowledge*. New York: Harper & Row.

Rosen, R. (1986). Causal structures in brains and machines. *International Journal of General Systems, 12*, 107–126.

Schuster, P. (1994). Extended molecular evolutionary biology: Artificial life bridging the gap between chemistry and biology. *Artificial Life, 1*, 39–60.

Toffoli, T. (1982). Physics and computation. *International Journal of Theoretical Physics, 21*, 165–175.

Turing, A. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, 42*, 230–265.

von Neumann, J. (1955). *Mathematical foundations of quantum mechanics*. Princeton: Princeton University Press. Chapter VI.

von Neumann, J. (1966). In A. Burks (Ed.), *The theory of self-reproducing automata*. Urbana: University of Illinois Press.

von Weizsäcker, C. F. (1973). Probability and quantum mechanics. *British Journal for the Philosophy of Science, 24*, 321–337.

Wheeler, J. A. (1990). Information, physics, quantum: The search for links. In W. H. Zurek (Ed.), *Complexity, entropy, and the physics of information*. Redwood City: Addison-Wesley.

Wheeler, J., & Zurek, W. (1983). *Quantum theory and measurement*. Princeton: Princeton University Press.

Whitehead, A. N. (1927). *Symbolism: Its meaning and effect*. New York: Macmillan.

Whitley, D., & Hanson, T. (1989). Optimizing neural networks using faster, more accurate genetic search. In *Proceedings of the third international conference on GAs* (pp. 391–396). Burlington, MA: Morgan Kauffman.

Wigner, E. P. (1964). Events, laws, and invariance principles. *Science, 145*, 995–999.

Wilber, K. (1985). *Quantum questions*. Boston: Shambala.

Zuker, M. (1989). On finding all suboptimal folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research, 9*, 133.

Zurek, W. H. (1990). *Complexity, entropy, and the physics of information*. Redwood City: Addison Wesley.

15 The Problem of Observables in Models of Biological Organizations

Abstract Physical theories have come to be the epitome of models for all levels of complexity, including living organizations, even though these theories are observer-independent and the observables in physical theories are relatively simple compared with those of biological and social models. In spite of this relative simplicity of observables, experimental physicists must spend most of their efforts designing and constructing complex devices to ground these observables by measurements. Theoretical physicists also invent non-observable constructs, and spend much of their efforts formalizing these imaginative constructs in syntactically unambiguous mathematical laws. For different levels of organization physicists choose different observables and laws, but at all levels the structure of physical theories sharply separates the observables from the constructs, and the measurement of observables from the computation of laws. Biological and social models are generally not observer-independent, nor do they separate their observables and constructs so clearly. I discuss the inadequacy of the physical model paradigm for modeling organizations that are complex enough to themselves be observers and modelers of their world. I explain why concurrent, distributed networks now used to model cognitive activity are a more appropriate paradigm for strongly interconnected, observer-dependent living organizations.

15.1 Introduction

The concepts and methodologies of physics have provided powerful tools that are useful for modeling some domains of biological and social systems. However, all models have limited domains, and the classical question always arises: *Are the*

Reprinted from *Evolution, Order, and Complexity*, Elias L. Khalil and Kenneth E. Boulding, Eds., London: Routledge, 1996, pp. 249–264.

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7,
DOI 10.1007/978-94-007-5161-3_16, © Springer Science+Business Media Dordrecht 2012

methodologies of physics adequate to model all the domains that characterize living organizations? I will discuss a domain where they are not. This is the domain of epistemic processes like observation, classification, recognition, and measurement that existing physical laws exclude in principle. Physical laws are universal in the sense that laws must hold at all places and at all times, and consequently laws must be independent of the state of the observer. This also means that observers have *no control* over the laws—laws are inexorable. In contrast, living organisms are characterized by *networks of controls*—genetic networks control enzyme synthesis, and enzyme networks control replication and gene expression, all metabolism and growth form networks of controls, and of course the nervous system is a many-level network of controls.

Physical laws do not usefully describe measurement and control functions. Of course measuring and control devices are material structures that obey laws and therefore a law-based reductionist description of these devices is certainly possible, at least in principle. However, such a detailed law-based description cannot derive or reveal the significant *functions* of these devices. This is even the case within physical theory itself where the choice of observables and the process of measurement are not derivable from physical laws.

The modern computer is also designed to be universal, but not in the same sense as laws are universal. A computer is universal by virtue of its independence from physical laws. Of course computers obey laws, but they are not run by laws, they are run by programs. Computers may be programmed to simulate laws, or controls or anything else. Anyone who wishes to describe biological behavior by detailed physical laws will find computers essential. However, the internal operations of a programmable computer are totally syntactic, isolated from physical analogs and any observational or measurement functions.

I begin by explaining in more detail why the universality of physical laws has an entirely different meaning from the universality of computation. Next, I discuss how our concepts of observation and measurement differ in physical and biological systems. I then give my reasons for favoring concurrent, distributed networks now used mostly for brain models, as necessary tools for modeling the epistemic functions and the complex, autonomous observables that have evolved in biological organizations. Such autonomous network models are also needed for a theory of observables in physics.

15.2 The Nature of Observables in Physical Theory

The domain of physics is restricted to those laws that are expressible in a mathematical formalism, and that satisfy invariance principles that keep the laws independent of the state of the observer. Classical physics studies the laws of universal and inexorable events that we feel "could not have happened any other way," as Wigner expresses it. Quantum theory has lost this classical determinism, but is even more structured by invariance or symmetry principles that assure the minimum influence of the observer on the formalism and the results of measurements. Quantum theory also differs from classical theory in the way the observables enter the formalism, but in both theories there is in every application a clear distinction between the constructs that are treated as measurable observables and those that are not.

This artificial separation of observable and imaginary constructs is embodied in the practice of physics by keeping the laws (expressed in a formal symbol system) and initial conditions (obtained by measurements) operationally distinct. The laws are represented by the equations of motion that define the family of patterns that can be calculated, and the measurements provide the initial conditions that specify which pattern corresponds to the world (e.g., Wigner 1964). Thus, the actual values of observable constructs can be grounded only by measurements, while the values of non-observable constructs can be obtained only by calculations. In practice, all scientists following this paradigm know without ambiguity whether at any moment they are making measurements or doing calculations.

Ideally, this separation guarantees that all the relevant contacts between the formal, syntactical aspects of the model and the world occur *only* during the measurement process. That is, the model is related to the world only through explicit observables. All other aspects of the model are essentially imaginative constructs that, beyond generating predictions, appear to be ontologically moot, and are justified only by highly informal metaphysical interpretations, and by aesthetic values such as simplicity, coherence and elegance.

One of the first statements of this idealized view of formal theories was given in 1894 by Hertz in the Introduction to his *Principles of Mechanics* (1956):

> We form for ourselves images or symbols of external objects; and the form which we give them is such that the necessary consequents of the images in thought are always the images of the necessary consequents in nature of the things pictured… With the things themselves they are in conformity in *one* important respect, namely, in satisfying the above requirement. For our purpose it is not necessary that they be in conformity with the things in any other respect whatever. As a matter of fact, we do not know, nor have we any means of knowing, whether our conceptions of things are in conformity with them in any other than this *one* fundamental respect.

By “one fundamental respect” Hertz did not mean that only one model was possible, since he stated that, “One image may be more suitable for one purpose, another for another.” His point was only that whatever image we use, the image has no verifiable truth value except at the perceived or measured “consequents”.

This principle that establishes a sharp cut between the world and any formal or syntactic model of the world also implies that any two formalisms that give the same measurable predictions are operationally equivalent. Two well-known examples of formally equivalent, but conceptually distinct models are the variational view and differential-equation-initial-condition view of trajectories, and the Schrödinger and Heisenberg descriptions of quantum mechanics. Of course, the conceptual and aesthetic differences between these formalisms, as in all formalisms, play an essential, role in our choice of theory (e.g., Polanyi 1958). Quantum theory is exceptional because its non-observable constructs are so completely imaginary that they appear to have no common-sense interpretation. This contributes to the view, as expressed by Wheeler (1982) that, “No elementary quantum phenomenon is a phenomenon until it is a recorded phenomenon” (i.e., the results of a measurement).

15.3 The Nature of Observation in Living Organisms

Organisms have evolved and learned to recognize patterns by direct sensing or perception that are far more numerous and complex than physical observables. These natural observables are selected because of their survival value, which often means that patterns must map to immediate actions rather than computations. Consequently, the organism’s measurement process is normally a *classification* of patterns that allow immediate decisions rather than the quantitative values necessary for computation. There are, of course some elementary observables, like length, time, force, and temperature that have quantitative physical measurement procedures that naturally correlate with our senses. These direct perceptions are usually felt to be closer to what we call reality than other physical concepts like potentials, vector fields, spins, and wave functions. This is a reasonable and useful belief for everyday survival, but nevertheless from the physicist’s point of view it remains only a metaphysical belief, since “reality” is not an observable of any physical theory. Fortunately, the predictive value of physical theory does not depend on which metaphysical view of reality we may prefer.

For many years I have used generalized *epistemic operations* like observation, detection, recognition, measurement, and control as the essential type of function that distinguishes living from non-living organizations (Pattee 1967, 1972, 1982). Survival requires the discovery of the significant patterns in the environment that can improve the organism’s control strategies. This recognition-control behavior is required for survival at all levels from genes to brains, and from families to societies. On the other hand, as I discussed above, physical theory while depending on well-defined observables, says nothing about how the significant patterns are discovered or selected, nor does the concept of control play any role in describing physical laws. Furthermore, the sharp separation between laws and measurements required by the universality of physical theory is not a characteristic of natural biological recognition-control processes.

My use of words like “observe” and “recognize” here must be generalized beyond their contextual usages in the several disciplines with which they are normally associated. Since I need to discuss the *evolution* of these functions, I always try to define the simplest cases. For example, I mean by an observer any system that recognizes patterns that subsequently are used to control actions. There are no special words or concepts for the most primitive cases. What is the simplest recognition process? What is the simplest control process? What is the simplest model?

These concepts are normally used at the cognitive level where we speak of recognizing patterns and performing controlled actions based on an intervening model. However, this same language can also be used at the molecular level to describe enzyme catalysis. The enzyme binds (recognizes) its specific substrate and catalyzes (controls) a specific reaction. There is no intervening model in this case, unless it is the structure of the enzyme molecule itself. My use of recognition and control at the enzyme level is not metaphorical, or at least no more metaphorical that its use at the cognitive level. In both cognition and catalysis the essential requirement for recognition-action or measurement-control function is that the mapping from pattern to action is arbitrary, or gratuitous, in Monod's (1971) sense. Any pattern can be coupled to any action depending on the structure of the cognitive model or the enzyme. If this were not the case, that is, if the measurement-control relation "could not be otherwise" then it could be described as a law.

Reductionists will of course claim that there is a valid physical description in terms of laws of any imaginable recognition, measurement, or control device, especially at the enzyme level. They would say that the apparent arbitrariness is simply a lack of a complete theory. This is a half-truth. While a lawful description is certainly possible in principle, no such description can define its own initial conditions. Therefore, as von Neumann (1955) pointed out long ago, such a description in terms of laws alone cannot describe the necessary *function* of measurement, and only leads to an infinite regress of increasingly complex lawful descriptions until terminated at some stage by functional measurements that are *not* describable by laws.

15.4 The Nature of Measurement in Primitive Organisms

It is still reasonable to question the use of such high-level terms as "measuring device" and "model" for a structure as simple as an enzyme. In normal usage, both concepts imply a much richer context. I have argued that the simplest context that would allow the normal use of epistemic concepts like measurement and observer is an organization that can *construct* the measuring device and use the results for its survival. In other words, measurement is not distinguishable by the local behavior of any mechanism. To qualify as a measuring device it must have a *function*, and the most primitive concept of function implies improving fitness of an organism. Thus, *observation and measurement require an organization that (1) constructs the measuring device and (2) uses the results of the measurements for survival*. This requirement I have called the *semantic closure principle* (Pattee 1982, 1995). This provides an objective criterion for distinguishing measurements and observations from other physical interactions. Only organizations with this semantic closure property should be called *observers*. The cell is the simplest natural case of an observing system. According to this view, simple artifacts that we commonly call measuring devices, like rulers, pendulums, and thermometers, are not intrinsically measuring devices, but perform the measurement function only by virtue of their role in the semantic closure that involves a human constructor, user, and interpreter. Of course, the same semantic closure requirement holds for simple, artificial controls, like governors and thermostats.

There are many relatively simple biological recognition-action structures that might suggest a primitive kind of model. For example, seedlings detect gravity and light, and by converting these input observables to specific rates of growth they control their morphology. A physiologist might prefer to call such tropisms a stimulus-response action and reserve the concept of model for a more complex relation between recognition and action. A cybernetician, on the other hand, would claim that the seedling has a model of its world, however primitive (Ashby 1956). In higher organisms we can recognize the nervous system as the physiological structure with the primary function of mapping sensory inputs from various receptor organs to output actions of muscles, and we often restrict the idea of model to mappings or representations in the brain. However, in the context of the more or less gradual process of evolution we do not learn much about primitive necessities for function by looking only at highly evolved organisms. There is generally more explanatory power in studying the origin of functions. What are the minimum requirements for this modeling relation in organisms?

An engineering description of a modeling relation would include at least three functions:

1. Detection, recognition, or measurement that transforms a physical pattern into model inputs. In organisms these are usually called receptors or sensory organs;
2. the model itself that establishes the particular input/output relation; and
3. the effectors that are controlled by the output of the model, and that interact again with the physical environment.

This very general description might to apply to the seedling example as well as the brain, but how do we make these divisions? Are these three functions intrinsic within all organisms themselves, or are they simply a convenient and conventional partition based on the physical and engineering paradigms of a model?

The practical criterion for choosing where the world ends and the measurement begins is simple, since the boundaries of organisms are usually sharp. But how do we determine where the measurement *ends* and the model begins? How do we determine when the model ends and the control begins? Since evolution is a more or less gradual process, the simpler the organism the less clear these distinctions become, but even in organisms with cells that are highly differentiated physiologically, it is not obvious how detection, modeling and control should be functionally differentiated.

15.5 Some General Criteria for Measurement Processes

These questions might suggest that it is the gradualism of evolution that makes it difficult to establish boundaries between measurements, models, and controls. However, since these same questions have not been answered clearly even in the case of simple, artificially designed physical measurements, it is clear that gradualism is not the only problem. How we should separate what we see as only physical interactions of an organism from perception remains a classical and still controversial problem in psychology. How do we determine the nature of what is observed, that is, how do we distinguish illusions and apparitions that are also observables. When do the various physical stimuli on the receptors become a recognized pattern in the brain?

There are many suggestions on how to objectify the measurement process. One strong criterion used in physics is that the completion of a measurement occurs only when a result has been *recorded* (see quote of Wheeler 1982, above). A record implies some kind of memory, but there are many types of record and many types of memory with all degrees of permanence—e.g., chalk marks, the grains of a photographic plate, magnetic tape, short and long term memory in the brain, etc. This criterion does not apply easily to the example of seedling growth, since there is no obvious record of interactions with light or gravity.

A weaker and more general criterion for a measurement is that the completion of a measurement *changes the probability* of selective future events. This implies that the result of a measurement gives us new information that updates our calculations of the probability of events. Boltzmann was probably the first to relate entropy to missing information, and Szilard (1929) showed qualitatively how a measurement by an observer can decrease the entropy of the system under observation. Of course to satisfy the second law of thermodynamics the measuring device itself must have a corresponding increase in entropy, but it is not the system under observation.

Note that this criterion applies to seedling growth, since changing the rate of selected directions of growth amounts to a change of probabilities. This criterion also applies to enzyme catalysis that is defined as a statistical bias on rates of specific reactions. This broad criterion even appears to apply to natural selection that can also be defined as a statistical bias on the relative rates of survival of hereditary units (Williams 1966). However, natural selection has no measuring device, unless it is the entire ecosystem, and I would exclude that by the semantic closure principle, since the individual organism (or observer) cannot be said to construct its ecosystem.

This selective statistical bias criterion for measurement also distinguishes the semantic interpretation of information from Shannon and Weaver's (1949) syntactic definition in which measurement processes play no role. The use of Shannon information measure is entirely arbitrary or gratuitous in any physical system for which the concept of fitness or function has no meaning, or for which "useful" work is not defined, since in these cases information and entropy are formally interchangeable. However, information arising from measurement must be distinguished from system entropy, since it can be "usefully" applied as a statistical bias (control) of the system's behavior, as in the Maxwell demon case. Of course, this distinction will remain largely arbitrary until we have a more objective criterion for when a measurement is completed.

15.6 The Evolution of Non-observable Constructs

Fortunately we do not have to solve the measurement problem in order to understand other important conditions that make the results of measurement more effective for survival and adaptation. Evolution has gradually differentiated and improved all functional organs, but the receptor and effector organs that interact directly with the world have strict functional limits determined by natural laws. For example, the sensitivity of detectors is limited by noise and the forces of effectors by the strength of materials. Such physical limits are approached in the receptors and effectors of many species. Consequently, after approaching these physical limits of input/output transducers, the only significant adaptations left are those occurring in the nervous system.

We recognize two types of brain models, instinctive models, or those inherited primarily from genetic instructions, and acquired models based on individual learning. What is epistemologically significant about primitive instinctive models is that there is little meaning to the distinction between non-observable constructs and observables or between literal and metaphorical models. The only possible test of an instinctive model is natural selection, and natural selection cannot distinguish imaginary constructs from measured observables in the organisms' models. *The only evolutionary distinction that can be made between models is between their relative survival rates.*

By contrast, our learned cognitive models are evaluated by how well they perform more limited, local functions. Usually these functions are tacit or poorly defined, although clearly the choice of function and quality of the model cannot evade natural selection. Physical models are exceptional because the observer-free universality principle and the separation of laws and measurements are explicit. Man has acquired enormous power from these principles, but still it is not obvious that such power improves his chance of survival as a species. Of course we have no way of knowing at what stage of evolution imagination and myth became a component of models important for survival. Certainly it was a gradual process, as it is in infant intellectual development.

One of many possibilities is that imagination improved on strategies of deception as well as its detection. How could this occur? Since deception, as well as its recognition, is already built into primitive instinctive strategies, to improve on these instincts would require a model of deception that can take into account learned experiences about deception. Any great improvement on instinctive models of deception could not consist of only direct observables, since some forms of deception are based on missing observables. In other words, a good model of deception must be made up of non-observable constructs. The same strategy works for discovering laws, even if nature is not deceitful.

In any case, whatever the survival functions of primitive imagination, it is this ability to invent mythical images or non-observable constructs that has, along with larger memories and more rapid learning, so effectively enlarged the domain of models. This domain of imaginary structures is apparently endless, and has evolved through all levels of meaning from vague and illusory images, from primitive dance to natural language with its rich myths, metaphors, and fictions, and only very recently on an evolutionary scale, to the invention of number, mathematics and the precise syntactic structures and rules of formal symbol systems. Modern physical theory now more than ever consists largely of these imaginative mathematical constructs that conform to reality only in Hertz's "*one* fundamental respect" that the observables agree at the points of completed measurements.

15.7 Sequential vs. Coherent Hybrid Models

The short-term historical evidence would suggest that models that depend either too much on imagination or too much on direct observation and measurement do not survive as well as hybrid models that use the advantages of both. That is, purely imaginary or mythical models, although they may correctly predict some events, carry so many gratuitous fictitious concepts that they also incorrectly predict most other events. On the other hand, models based only on direct observation and measurement with literal, unimaginative interpretations will generate only local extrapolations, and miss the predictive and explanatory power of universal laws that are guided by unobservable logical and aesthetic principles. Of course myths and aesthetic principles are themselves undoubtedly influenced by primitive genetic models. (Casirer 1957; Piaget 1978).

What are the strategies for combining observables and non-observables in a hybrid model? As I have described above, physical theory is one extreme strategy that tries to formalize all non-observables so that they are invariant with respect to observers. This syntactical formalism of laws is semantically grounded only by discrete sequences of measurements performed at times or intervals that are necessarily unpredictable by the theory. One way to say this is that the formal structures of laws form *coherent, rate-dependent* dynamics established by a universal time parameter. The computation of laws requires discrete symbols and *rate-independent* discrete steps. It is the measurement process that maps continuous dynamics into discrete, rate-independent (recorded) symbols. The invariance principles also demand that the measurement event cannot be described by this coherent dynamics, since the time of measurement is the observer's choice. Measurement in physical theory is therefore *incoherent*, and must be a non-dynamical, sequential process (von Neumann 1955).

A second less principled strategy of hybrid models is characteristic of the engineering and cybernetic disciplines where the purpose of models is to *control* the behavior of artificial systems according to intelligently designed functional specifications. In engineering, finding physical laws is not the problem. The laws are assumed, and the problem is to find machines that constrain these laws to fulfill the required function. In these applications, observables are chosen according to their value as *control* parameters for attaining the specified function. Unlike measurements in physics which are discrete and relatively infrequent, engineering measurements used for control are often made continuously, as in servomechanisms. Measurement in this case becomes a coherent part of the dynamics. Since this type of control is also characteristic of living organisms, they appear to be much more amenable to description by engineering models than by fundamental physical models. However, engineering models also partition the measurement, model, and control functions in constructing machines. Also, as Polanyi (1968) has emphasized, all machines are designed only as prosthetic devices by organisms with brains, and therefore engineering models seldom have any explanatory power for the origin of observables.

15.8 Computer Models

The most recent addition to scientific modeling tools is the programmable computer. Computers can represent complex behavior that in some ways is beyond the capacity of the brain, and therefore the significance of computer models is sometimes difficult to evaluate. A new appreciation of nonlinear dynamical behavior (strange attractors, chaos, cellular automata, fractals, etc.) has resulted from computer experiments that have caused us to reconsider our views of determinism and random behavior as well as our choice of significant observables in complex systems. As a result of these computational discoveries, we are more willing to consider the possibility that much detailed biological and social behavior is in principle unpredictable, not because of ignorance or noise, but because of chaotic dynamics.

However, the discovery of new observables remains outside the domain of computer programs, and requires the pattern-recognizing skills of an intelligent observer. Such "empirical" computations are more accurately interpreted as analogs or artificial worlds in which an outside observer can discover predictable statistical relations and patterns of behavior that would otherwise be undetectable or unrecognizable. All programmable computational processes require that the inputs are already in symbolic form (i.e., all observables have been chosen, and all measurements have been completed), so that the measurement problem does not arise. Therefore the programmable computer itself cannot be classed as a hybrid model, since the distinction of observable and non-observable, if it is made at all, must be made by an interpreter outside the computer. At the cost of having no intrinsic contact with the world, programmable computers are symbolically universal. Programmable computers can therefore simulate anything that can be defined symbolically, but insofar as a computation is purely syntactic, its operation should not be called a *theory* of anything. Of course the same is true of any purely formal mathematical symbol system.

15.9 Limits of Universality in Physical and Computational Models

Both physical laws and computation are described as universal, which often is used as an argument for reductionism—everything can be described by laws and simulated by computation. However, "universal" has a different meaning in each case, and both are strictly limited. In physical theory, the concept of physical law is universal only with respect to a set of precisely defined observables (e.g., Eddington 1928). All candidates for a law must in principle apply everywhere and at all times, i.e., under all conditions of observation. This is not a demonstrable fact, but an epistemological requirement to distinguish what we call objective reality from an individual's local perceptions. The reductionist's claim that all possible observables are derivable from the atomic observables of physical laws remains a metaphysical faith, since not even in physics are observables derivable from the laws.

Computational universality has nothing to do with observables in the physicist's sense, but is a purely formal statement about equivalent processes of unambiguously rewriting strings of discrete symbols. The term universal arose from Turing's discovery of his universal machine that could imitate the computation of any definable Turing machine. This formal concept of "machine" was extended by proofs to other types of symbol-string rewriting rules equivalent to Turing's universal machine, and by the informal (nonprovable) Church-Turing thesis that the concept of effective computability did not extend beyond these equivalent formalisms.

Both of these "universal" methods have certain other intrinsic limitations that are well-known, but not always interpreted in the same way. In physics there are the uncertainty relations that assume quantum mechanics is a complete description (i.e., no hidden variables), but in which predictive determinism is impossible because of limitations on the precision of simultaneous measurement of conjugate observables (e.g., position and momentum). A kind of converse limit occurs in formal systems that assume determinism (i.e., effective computability), but in which completeness is proven impossible (e.g., consistency is not provable, noncomputable functions exist).

These limitations have been used in arguments to escape from the classical determinism of laws and computation that to some appear necessary for modeling emergence, creativity, and any novel behavior characteristic of life. The oldest attempts to escape from classical physical determinism invoke the probabilistic interpretation of quantum mechanics. Gödel incompleteness has been used as an escape from syntactic determinism, and as evidence that minds are not machines. More recently the idea of symmetry-breaking instabilities has been added to escape theories (e.g., Prigogine and Stengers 1984), and the newest escape from determinism is chaotic dynamics (e.g., Ford 1989). There are many other ideas on the inadequacy of the physical and computational paradigms for models of life that are too lengthy to be described here (e.g., Lockwood 1989; Penrose 1989; Rosen 1991; Cariani 1992).

These are all profound limitations, and yet if they could in some way be evaded or overcome I do not see that the problem of discovery of observables, and the arbitrariness of measurement would be resolved. According to my view here, ideas like determinism and chance fall within the enormous class of non-observable constructs that may be conceptually useful for some models and only confusing for others. My point is that *whether one or another non-observable construct is appropriate for a model depends on the observables chosen for the model*. In other words, it can only be of metaphysical concern if one model of a system uses deterministic constructs and another model of the "same" system uses stochastic constructs as long as the observables of the two models are different. This view is not new. Such complementary models are well-known in the *reversible* mechanics of particles with position and momentum observables, and *irreversible* thermodynamical descriptions of a population of particles with statistical averages as observables. Of course in quantum mechanics the necessity of such incompatible constructs are the basis of Bohr's (1927) complementarity principle.

15.10 Concurrent Distributed Network Models

Distributed networks or connectionist machines perform *classifications*, and that is one necessary condition for an organism to discover an observable. What does classification mean? Logically a classification requires a many-to-one mapping, and to realize this in a physical system means that the classes correspond to statistical stationary states reachable from many initial conditions. This idea that a local statistical equilibrium could correspond to a class of input patterns was a key motivation for many distributed network models of cognition (e.g., Anderson and Rosenfeld 1988).

In addition to their potential as models of cognitive activity, network models apply to a wide variety of biological and social models. Living organizations are highly interconnected as networks at many levels from the metabolic networks of enzyme catalyzed reactions, multicellular organisms and social groupings, to ecological networks. All of evolution has occurred in such metabolic and ecological networks. Kauffman (1991) and others have shown how such network models can be instructive from the molecular level to evolution even though we do not know the detailed connectivity of such networks, since even random networks exhibit robust generic behaviors that accomplish a type of statistical self-organization that is necessary for the appearance of autonomous classifications.

Some computationalists try to interpret networks as only another architecture that realizes the logical computation of a universal Turing machine. This is based on the invention of codings that show formal computational equivalence, but this view ignores the fact that new codes and different architectures can generate new patterns and statistical observables that have no meaning at the formal symbol-rewriting level. A more physical interpretation of networks is that they represent an analog statistical dynamical system that, like all dynamical systems, happens to be representable by programming a universal computer (e.g., Smolensky 1988).

Ironically, the idea of logical computation was initially motivated by the desire to understand thought rather than physical laws. Descartes and Leibniz believed that computation could represent thought itself, and many of the founders of the modern programmable computer were first motivated by their interest in logic and the brain. McCulloch and Pitts, who knew much neurophysiology, were trying to see if brains could be modelled as a network of logical elements; so were von Neumann and Weiner. The later, well-known contributions to computer engineering by these men was motivated by the practical coding and ballistics problems of World War II, rather than attempts to model the brain (Weiner 1948).

In spite of these roots of computer theory in attempts to model brains as networks, and von Neumann's (1955) cautions that the brain was "not digital, but statistical," and that "the language of the brain [is] not the language of mathematics," the next generation that founded the field of artificial intelligence ignored these original, biologically based network concepts, and postulated that all cognitive activity was computational in the logical or rule-based sense. The fundamental physical concepts of observables and measurements were completely excised from the domain of cognitive models by the *The Physical Symbol System Hypothesis* (Simon and Newell 1976) that defined intelligent action as programmable, symbolic computation. Logical programmed computation became the dominant model for thought for over two decades, largely eclipsing network models. Logical programming was of course supported by the rapid growth of sequentially programmed computer architectures and hardware technologies.

The reasons for the rather sudden revitalization of network models are not obvious, and involve many complex people and events (e.g., Papert 1988; Dreyfus and Dreyfus 1988; Rumelhart and McClelland 1986). However it is clear that several fundamental papers that influenced network research were written by physicists (Cooper 1973; Little and Shaw 1975; Hopfield 1982) who saw networks as *analogs* of physical systems, not as logical computers. The fact that network dynamics is computed by discrete symbol manipulation is a matter of technical speed and precision, just as it is in calculating the orbits of celestial bodies. Their discovery of interesting observables in network patterns occurred by the same cognitive processes that interesting observables are discovered in the real world.

15.11 The Complementarity of Physical and Network Models

Distributed networks are a promising model for exploring how new observables are discovered because they can autonomously classify patterns. However, an observable is more than a classification. If we say that observables are that part of a model of the world that can be measured or recognized, we are slipping into the dichotomy of physical theory that separates the measurement from the model. Recall that the principle requiring this dichotomy was that physical models be *observer-independent*, or objective. But objectivity is seldom a biological requirement for survival. In fact our individuality is based on our separate memories and models that are uniquely *observer-dependent* or subjective.

We can imagine two ideal types of model that are complementary, one type for discovery of universal laws, and the other type for controlling individual actions. In Ryle's terms, one type of model is for "knowing that" and another type is for "knowing how". It is only because of the epistemological requirements for objectivity and universality that the ideal discovery model requires observer-independence. This in turn requires strict separation of the *actions* of individual observers from the dynamics of universal laws. Observations and measurements are such individual actions.

In contrast, the ideal control model benefits the individual and has no requirements for objectivity or universality. Therefore it need not separate measurements, models, and actions. These categories may not even be appropriate for describing the behavior of a brain's distributed networks that continuously modify their complex inputs and generate output patterns or equilibria. How can one say when measurement ends and modeling begins when these functions are concurrent and distributed? How can observables be distinguished from imaginary constructs if neither is localized in the network?

And yet it is clear that some activities in the brain allow us to invent observables and non-observable constructs, to imagine the sharp classifications that allow physicists to separate laws and measurements, the observed from the observer, and that allow mathematicians to imagine the crisp, formal symbol systems with which we calculate. Physical theory attends explicitly only to the universal laws, but cannot explain the discovery of observables and the measurement process without also attending to models of the observer. Biological theory will need to attend to distributed network models to explain emergence of observables in evolution and learning. Network models must also lead to classifications that support objective models. This suggests that a complete description of either living or inanimate behavior requires both types of model—observer-independent and observer-dependent. These models are complementary in Bohr's sense that neither type of model is derivable from or reducible to the other, and both are necessary for understanding physical laws and life.

References

Anderson, J. A., & Rosenfeld, E. (1988). *Neurocomputing: Foundations of research*. Cambridge, MA: MIT Press.

Ashby, R. (1956). *An introduction to cybernetics*. London: Chapman & Hall, Ltd.

Bohr, N. (1927). The 1927 Como Lecture, reprinted in *Nature 121*, 580.

Cariani, P. (1992). Emergence and artificial life. In C. Langton, C. Taylor, J. Farmer, & S. Rasmussen (Eds.), *Artificial life* (Vol. II, pp. 775–797). Redwood City: Addison-Wesley.

Casirer, E. (1957). *The philosophy of symbolic forms* (The phenomenology of knowledge, Vol. 3, pp. 775–797). New Haven: Yale University Press.

Cooper, L. N. (1973). A possible organization of animal memory and learning. In B. Lundquist & S. Lundquist (Eds.), *Proceedings of the Nobel symposium on collective properties of physical systems*. New York: Academic.

Dreyfus, H. L., & Dreyfus, S. E. (1988). Making a mind versus modeling the brain: Artificial intelligence back at a branchpoint. *Daedalus, 117*, 15–43.

Eddington, A. S. (1928). *The nature of the physical world* (p. 260). New York: Macmillan.

Ford, J. (1989). What is chaos that we should be mindful of it? In P. Davies (Ed.), *The new physics* (pp. 348–371). Cambridge: Cambridge University Press.

Hertz, H. (1956). Introduction. In *The principles of mechanics* (pp. 1–2). New York: Dover (Orig. German ed., 1894).

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of National Academy of Science, 79*, 2554–2558.

Kauffman, S. (1991). *Origins of order*. London: Oxford University Press.

Little, W. A., & Shaw, G. L. (1975). A statistical theory of short and long term memory. *Behavioral Biology, 14*, 115–133.

Lockwood, M. (1989). *Mind, brain & quantum*. Oxford: Blackwell, Ltd.

Monod, J. (1971). *Chance and necessity*. New York: Alfred Knopf.

Newell, A., & Simon, H. (1976). Computer science as an empirical enquiry. *Communications of the Association for Computing Machinery, 19*, 113–126.

Papert, S. (1988). One AI or many? *Daedalus, 117*, 1–14.

Pattee, H. H. (1967). Quantum mechanics, heredity, and the origin of life. *Journal of Theoretical Biology, 17*, 410–420.

Pattee, H. H. (1972). Physical problems of decision-making constraints. *International Journal of Neuroscience, 3*, 99–106.

Pattee, H. H. (1982). Cell psychology: An evolutionary approach to the symbol-matter problem. *Cognition and Brain Theory, 5*(4), 325–341.

Pattee, H. H. (1995). Evolving self-reference: Matter, symbols, and semantic closure. *Communication and Cognition – Artificial Intelligence, 12*(1–2), 9–28.

Penrose, R. (1989). *The emperor's new mind*. Oxford/New York: Oxford University Press.

Piaget, J. (1978). *Behavior and evolution*. New York: Pantheon.

Polanyi, M. (1958). *Personal knowledge*. New York: Harper & Row.

Polanyi, M. (1968). Life's irreducible structure. *Science, 113*, 1308–1312.

Prigogine, I., & Stengers, I. (1984). *Order out of chaos*. New York: Bantam Books.

Rosen, R. (1991). *Life itself*. New York: Columbia University Press.

Rumelhart, D., & McClelland, J. (1986). *Parallel distributed processing*. Cambridge, MA: MIT Press. Chapter 1.

Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavior and Brain Sciences, 11*, 1–74, and continuing commentary, *ibid.* 13, 399–412 (1990).

Szilard, L. (1929). Uber die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen. *Zeitschirft fur Physik, 53*, 840–853.

von Neumann, J. (1955). *The mathematical foundations of quantum mechanics*. Princeton: Princeton University Press.

Weiner, N. (1948). *Cybernetics*. Cambridge, MA: MIT Press.

Wheeler, J. A. (1982). Bohr, Einstein, and the strange lesson of the quantum. In R. Elvee (Ed.), *Mind and nature* (p. 9). New York: Harper & Row.

Wigner, E. (1964). Events, laws, and invariance principles. *Science, 145*, 995–999.

Williams, G. C. (1966). *Adaptation and natural selection* (p. 22). Princeton, NJ: Princeton University Press.

16 Causation, Control, and the Evolution of Complexity

16.1 Is Causation a Useful Concept?

It is not obvious that the concept of causation, in any of its many forms, has ever played a necessary role in the discovery of the laws of nature. Causation has a tortuous philosophical literature with no consensus in sight (e.g., Hart and Honoré 1958; Bunge 1959; Taylor 1972), and modern physics has little interest in the concept. Nevertheless, causation is so ingrained in both the syntax and semantics of our natural language that we usually feel that events are somehow causally explained by almost any grammatically correct declarative statement that relates a noun and a verb phrase to the event: Why did the ball roll? Because John kicked the ball. Why did the ball bounce? Because the ball hit the post. In Aristotelian terms, the verb is a form of efficient cause, and either the subject or object can act as a material cause. If the subject happens to have a large brain we may also attribute a formal, teleological, or intentional cause to the event: Why did John kick the ball? Because John wanted a goal. As a child we figure out that these linguistic forms are transitive and always lead to a vicious circle or an infinite regress, but we are usually told that it is rude to continue to ask, "Why?" when presented with one proximal cause. The major weakness of the concept of causation is this Whorfian dependence on natural language. Thus, the richness and ambiguity of causal forms arises more from the richness and ambiguities of language than from any empirical necessity or from natural laws.

Reprinted from *Downward Causation: Minds, Bodies, Matter*, P. B. Anderson, C. Emmeche, N. O. Finnemann, P. V. Christiansen, Eds. Aarhus University Press, 2000, pp. 63–77.

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7, DOI 10.1007/978-94-007-5161-3_17,

16.2 Naive Causation Requires a Direction in Time

This naive concept of causation is formed from our perception of certain sequences of events. One condition is temporal antisymmetry. That is, when we say an event B is caused by an event A, it must be the case that A occurred before B. If temporal order were reversed any cause and effect relation would also be reversed, although some philosophers have questioned this assertion (e.g., Dummett 1964). The concept of causation therefore presupposes a model of time, usually a tacit model. Our everyday concept of time is directed in one dimension, and so we ascribe causation to events that can be decomposed into simple strings of ordered events or actions. The high-dimensional and diffuse concurrent influences that are ubiquitous are seldom viewed as causes. However, like the concept of time, the meaning of causation does not easily lend itself to deeper analysis. When we try to define more precisely the concepts of time and causation we find they are entirely context or model dependent. Furthermore, these concepts are often not consistent between contexts and levels. To make matters worse, they usually appear as irreducibly primitive concepts at all levels.

16.3 Causation Is Gratuitous in Modern Physics

The Newtonian paradigm of state-determined rate laws derived from a scalar time variable and explicit forces only strengthens the naive concept of one-dimensional, focal causation. Reductionists take the microscopic physical laws as the ultimate source of order. At this lowest level, causation was classically associated with the concept of force. According to one statement of Newton's law a force is the cause of objects changing their motion. The concept of force can also be interpreted in many ways, but in practice most physical models are of systems with a very small number of forces, or more precisely, of systems where the equations of motion can be easily integrated or computationally iterated. However, in the case of the famous n-body problem ($n>2$) that is generally nonintegrable, the forces are so interdependent that no focal causes exist. The motion of one body in an n-body model might be seen as a case of downward causation, but this does not add anything to our understanding of the fundamental problem.

The fundamental problem is that the microscopic equations of physics are time symmetric and therefore conceptually reversible. Consequently the irreversible concept of causation is not formally supportable by microphysical laws, and if it is used at all it is a purely subjective linguistic interpretation of the laws. Hertz (1894) argued that even the concept of force was unnecessary. This does not mean that the concepts of cause and force should be eliminated, because we cannot escape the use of natural language even in our use of formal models. We still interpret some variables in the rate-of-change laws as forces, but formally these dynamical equations define only an invertible mapping on a state space. Because of this time symmetry, systems described by such reversible dynamics cannot formally (syntactically) generate intrinsically irreversible properties such as measurement, records, memories, controls, or causes. Furthermore, as Bridgman (1964) pointed out, “The mathematical concept of time appears to be particularly remote from the time of experience.” Consequently, no concept of causation, especially downward causation, can have much fundamental explanatory value at the level of microscopic physical laws.

16.4 Do Statistical Laws Give a Direction to Time?

The answer to this question is still controversial. It is a near tautology to state that on the average the more likely events will occur sooner than the less likely events. In the more precise form of the second law of thermodynamics this is still a useful near-tautology. Here the word “sooner” appears to give time a direction, as does the second law’s increasing entropy or disorder with time in an isolated system. But on careful thought we see that sooner and later are concepts that presuppose a direction of time. This statement, and the second law, would still be true if time were reversed since sooner and later would also be reversed. Assuming an isolated system with less than maximum entropy, the plot of entropy vs. time would show increasing entropy in both directions of time without favoring either direction (e.g., Tolman 1950). Nevertheless, it has been argued on many grounds that the observer’s psychological time must be consistent with the second law, and furthermore, using the weak anthropic principle, both must correspond to the cosmological arrow of time (e.g., Hawking 1988).

What is important to recognize is that the concepts of causation have completely different meanings in statistical models and in deterministic models. A reductionist will assume that cause refers to events in a lower level model. That is, if we ask what is the cause of temperature, the reductionist will answer that it is caused by molecules exchanging their kinetic energy by collisions. But notice that the measurement of temperature is practical only because measuring devices effectively average this exchange without requiring measurement of detailed initial conditions of all the molecules. Averaging is not part of the microscopic model but is a statistical process of a higher level model. A deterministic microscopic model cannot cause an average to be an observable. There is also the model of flipping a coin. Here the reductionist will again say that it is the detailed initial conditions that determine the result, but in this case precise enough measurement of initial conditions is not practical, and therefore flipping a coin is modeled as a random event.

16.5 Measurement Gives a Direction to Time

Many people are satisfied by the reductionist’s detailed “causes” and feel that these microscopic models have explained the macroscopic observations. However, a skeptic will observe that averages, coin equilibria, dissipation, measurement, and all other irreversible or stochastic events cannot be derived from reversible, deterministic models, and therefore cannot be adequately reduced to, or explained by, such models (e.g., Coveney and Highfield 1991). In the two examples above, what forces an asymmetric direction of time in our models is not the microscopic behavior of the system, but the measurement process. In the case of temperature, the irreversible process of averaging is done by the measuring device, the thermometer, not the reversible dynamics of the molecules of the system being measured. For the same reason, the macroscopic observables of heads or tails of a coin appear only after the reversible dynamics of the coin have been dissipated and the coin has come irreversibly to rest. Dissipation here simply means that the useful details of the motion have become unmeasurable.

From this line of argument we conclude that our concepts of the direction of time and hence our concepts of causation arise from our being observers of events, not from the events themselves. Consequently concepts of causation are subjective in so far as they cannot be separated from the observer's choice of observables and the choice of measuring devices. According to this model one might be tempted to say that it is the observer who causes a direction to time, not physical laws, but this would overstate the causal powers of the observer. Physical explanations require an epistemic cut between the knower and the known, and a model of the observer on one side of the cut makes no sense without the complementary model of the laws of the observed system on the other side (e.g., von Neumann 1955).

16.6 Universal Causes Are Not Explanatory

The reductionist's answers above are examples of universal causes. It is a metaphysical precondition for physical laws that they must hold everywhere for all observers. Laws are inexorable. That is, we expect every event at any level of complexity to satisfy these laws no matter what higher level observables may also be needed for a useful model. Therefore, just as it is correct to say that the temperature in this room is caused by atoms following the laws of physics, it is equally correct to say that the cause of my writing this paper is the atoms of my brain following the laws of physics. But since such statements hold in all conceivable cases they give no clue to the level of observables necessary for a useful model in each case. It is only our familiarity with this linguistic form that often leads us to accept uncritically such universal causes as explanations.

16.7 Complementary Models Require Complementary Causes

We know from the two fundamental levels of physical models, the microscopic laws and the statistical laws, that it is a wasteful exercise to try to abstract away the differences between these models since they are complementary. I am using complementary here in Boltzmann's and Bohr's sense of logical irreducibility. That is, complementary models are formally incompatible but both necessary. One model cannot be derived from, or reduced to, the other. Chance cannot be derived from necessity, nor necessity from chance, but both concepts are necessary. In his essay on dynamical and statistical laws, Planck (1960) emphasizes this point: "For it is clear to everybody that there must be an unfathomable gulf between a probability, however small, and an absolute impossibility. (…) Thus dynamics and statistics cannot be regarded as interrelated." Weyl (1949) agrees: " (…) we cannot help recognizing the statistical concepts, besides those appertaining to strict laws, as truly original." And similarly, von Neumann (1955) in his discussion of measurement says: "In other words, we admit: Probability logics cannot be reduced to strict logics." It is for this reason that our concept of a deterministic cause is completely different from our concept of a statistical cause. Determinism and chance arise from two formally complementary models of the world. We should also not waste time arguing whether the world itself is deterministic or stochastic since this is a metaphysical question that is not empirically decidable.

These examples show the extreme forms and model-dependencies of our many uses of causation. Notice that both complete determinism and complete chance can be invoked as causal "explanations" of events. These extreme forms of causation are often combined to describe what we see as emergent events that require new levels of description as in symmetry-breaking and dissipative structures in physical models (e.g., Anderson and Stein 1988), or what Crick called "frozen accidents" in biological models.

16.8 Useful Causation Requires Control

As I noted above, the use of causation at the level of physical laws is now considered as only a gratuitous manner of speech with no fundamental explanatory value. Naturally the question arises: At what level of organization does the concept of causation become useful? To explain my answer to this question let me first jump up several levels of complexity. Clearly it is valuable to know that malaria is not a disease produced by "bad air" but results from Plasmodium parasites that are transmitted by Anopheles mosquitoes. It is also valuable to know that the lack of vitamin C will result in scurvy. What more do we gain in these examples by saying that malaria is caused by a parasite and scurvy is caused by lack of vitamin C?

I believe the common, everyday meaning of the concept of causation is entirely pragmatic. In other words, we use the word cause for events that might be controllable. In the philosophical literature controllable is the equivalent of the idea of power. Bishop Berkeley thought it obvious that cause cannot be thought of apart from the idea of power (e.g., Taylor 1972). In other words, the value of the concept of causation lies in its identification of where our power and control can be effective. For example, while it is true that bacteria and mosquitoes follow the laws of physics, we do not usually say that malaria is caused by the laws of physics (the universal cause). That is because we can hope to control bacteria and mosquitoes, but not the laws of physics. When we say that the lack of vitamin C is a cause of scurvy, all we mean is that vitamin C controls scurvy. A fundamental understanding or explanation of malaria or scurvy is an entirely different type of problem.

Similarly, when we seek the cause of an accident, we are looking for those particular focal events over which we might have had some control. We are not interested in all those parallel, subsidiary conditions that were also necessary for the accident to occur but that we could not control, or did not wish to control. For example, when an aircraft crashes there are innumerable subsidiary but necessary conditions for the accident to occur. When we look for "the cause" of the accident we are not looking for these multitudes of necessary conditions, but for a focal event that, by itself, might have prevented the accident but maintained all other expected outcomes.

In our artificial technologies and in engineering practice we also think of causes in terms of control. For example, the electrical power that provides the light in my room is ultimately caused by nuclear fission in the sun that drives the water cycle and photosynthesis, or by nuclear fusion on earth. Many complex machines and complex power distribution systems are also necessary in the causal chain of events lighting my room. So why do I think that the cause of the light in my room is my turning the switch on the wall? Because that is where I have proximal, focal control, and also because switching is a simple act that is easy to model, as contrasted with the complexities of nuclear reactions and power distribution networks.

We view the causal aspects of all our machines in this way. We do not think of any very complex system or diffuse network of stochastic influences as a cause. This is one reason that downward causation is problematic. In other words, we think of causes in terms of the simplest proximal control structures in what would otherwise turn into an endless chain or network of concurrent, distributed causes. A computer is a useful modeling device because the simple, controllable steps of a program are the pragmatic cause of the computer's behavior. It is also significant that at the cultural level of jurisprudence it is only those causes that are focal, explicit, and believed to be controllable that are admissible in determining guilt or innocence. No jury will acquit by reason of downward causation.

16.9 The Origin of Control

The lack of any obvious explanatory power or utility of the concept of causation at the level of physical laws led to the question of what level of complexity causation does become useful. I supported the classical philosophical view that causation is a useful concept only when associated with power and control. This leads to the next question: At what level of organization does the concept of control become useful? The concept of control does not enter physical theory because it is the fundamental condition for physical laws that they describe only those relations between events which are invariant with respect to different observers, and consequently those relations between events over which the observer has no control.

At the least, control requires, in addition to the laws, some form of local, structural constraint on the lawful dynamics. Pragmatic control also requires some measure of utility. To say the riverbed controls the flow of the river is a gratuitous use of control since there is no utility, and the simpler term constraint serves just as well. Following the pragmatic requirement that concepts of causation and control must have some utility, I would say that utility makes sense only in terms of some form of fitness or function of a system that is separate from, but embedded in, an environment. Just as the concept of measurement requires an epistemic cut between the measuring device and the system being measured, so the concept of control requires an epistemic cut between the controller and the controlled.

Living organisms are the first natural level of organization where we know these concepts of functional control and fitness in an environment clearly make sense, and in fact are necessary for a useful model. Of course artifacts are also functional, but these are products of living organisms. While there must be intermediate levels of organization from which our present forms of life arose, the fact is that present life requires semiotic control by coded gene strings. There are many theories of self-organization that try to fill in these intermediate levels (e.g., Eigen and Schuster 1982; Nicolis and Prigogine 1989; Kauffman 1993; Langton 1989), but at present there exists an enormous gap between these statistical physics and artificial computer-life models and the complex, coded, semiotic control of life as we know it. It is arguable whether the concepts of causation and control are necessary or useful in these intermediate level models. Often the use of such high-level concepts of natural language to describe simple models obscures the real problem.

Why do most of us first think of the gene as the primary causal structure of the organism even though we know that some form of downward causation from the organism level is essential to control which genes are expressed? Again, one answer is that the gene's control activities are local, sequential, and relatively easy to model, as contrasted with the organism's downward control which is diffuse, parallel, and complex. However, there is a more fundamental reason: Genetic control is heritable—it is stored in a relatively simple, localized, semiotic memory that is easy to transmit. The organism's downward controls are not stored in memory, but are part of the time-dependent dynamics of the phenotype. Phenotypic dynamics are neither simple, localized, nor heritable.

16.10 Levels of Control Match Models of Causation

The pragmatic view of causation implies that different levels of causation will be associated with different levels of control. Downward causation is a difficult concept to define precisely because it describes the collective, concurrent, distributed behavior at the system level where control is usually impractical, rather than at the parts level where focal control is possible. Downward causation is ubiquitous and occurs continuously at all levels, but it is usually ignored simply because it is not under our control. For example, even in relatively simple artificial neural nets we know that collectively the hidden nodes exert downward control on the output. Yet while we have some control training at the level of the entire net we rarely know how to explicitly control at the level of individual hidden nodes.

In real-life the problem is much worse. In the real brain we may exert some control by drugs at the coarse level of awareness and moods, or somewhat finer control by brain surgery, but the firing of individual neurons is not controllable in any useful way. The same situation occurs at all levels, in ecosystems, social systems, economic systems, and even in systems that are designed to be controllable but that have grown excessively complex. Some catastrophic system failures, including cancer, aging, death, and species extinctions that might be viewed as a form of downward causation could just as well be described as loss of detailed control.

16.11 Evolution Requires Semiotic Control of Construction

This fundamental problem of how the dynamics of life maintains, or increases, its control of complexity while most nonliving dynamics tend to decay was one of Boltzmann's deepest concerns, but he found no satisfactory answer. The first hint of the answer was suggested by von Neumann (1966) in his discussion of complication and his theory of self-reproducing automata. Von Neumann was also motivated by the apparent conflict between structures that decay and structures that evolve. He focused on automata models, but it is clear that he had the contrast between thermodynamics and biological evolution in mind. He saw in Turing's universal automaton an example of a simple, fixed symbol system that could generate open-ended complexity. In order to translate this open-endedness to a physical system, von Neumann first postulated a universal constructor that could interpret symbolic descriptions. The universal constructor, like Turing's universal machine, was relatively simple, but the descriptions could grow indefinitely and consequently the resulting constructions could grow in complexity. The essential property of semiotic description is that it can be read in two ways: it can be read syntactically to be transmitted, and it can be read semantically to control construction.

Today we know in great detail how cells reproduce and evolve using this fundamental description-construction strategy. Over evolutionary time scales the cell's construction machinery (tRNA, aminoacyl synthetases, ribosomes) remains more or less constant, but the gene grows in length and the organism grows in complexity. This dependence of life on the separation of genotype and phenotype has been implicit in evolution theory since Darwin, but it is only recently that the adaptive power of genetic search in sequence space and its redundant mapping to structure has begun to be understood. This power has been discovered largely by empirical exploration of adaptive systems by computer models of maps from sequence space to structure space (e.g., Schuster 1994), and sequence space search using genetic algorithms (e.g., Holland 1992; Goldberg 1989). The combination of crossover and mutation has been shown to be surprisingly powerful for finding solutions of certain classes of problem that are otherwise intractable. It is not yet clear why genetic algorithms work well in some cases and not in others. The building-block hypothesis and schema theorem are part of the answer.

What is clear is that successful evolution depends on both the structure of the sequence space of the gene for efficient search, and how sequence space maps to function space by control of constructions (e.g., Conrad 1990). The details of this mapping from genetic description to physical rate dynamics is a difficult empirical problem, but the fundamental requirement for open-ended evolvability is the interdependence of the semiotic domain of the heritable genetic memory and the dynamic domain of construction and function.

16.12 Artificial Dynamics and Self-organization

It is now well known history how semiotic rule-based systems dominated artificial intelligence models until the rediscovery of the potential of nonlinear dynamics and concurrent, distributed network models. With the rediscovery of the adaptive power of networks, the study of nonlinear dynamic behavior has now largely replaced the rule-based symbolic models of artificial intelligence. In evolution theory there has also been a shift in interest toward dynamical models of self-organization as a non-exclusive alternative to the traditional heritable genetic variation and selection theory of evolution. The current controversy is over how much of evolution and development results from genetic control and natural selection and how much from self-organizing nonlinear dynamics. At the cognitive level the corresponding controversy is over how much of our thinking is the result of sequential semiotic rules and how much is the result of distributed, coherent neural dynamics.

These questions will not be resolved by either-or answers—first, because semiotics and dynamics must be intricately related at all levels of organization, precisely because it is this semiotic-dynamic interaction that is responsible for evolving levels, and second, because semiotic and dynamic models are complementary, both conceptually and formally. Conceptually dynamical models describe how events change in time. Since time is viewed as continuous and one-dimensional, nonrelativistic dynamical processes are conceptually viewed as concurrent, coherent, or parallel in time, no matter how many variables or other dimensions exist. Dynamical laws are state determined; we need only know the initial conditions; there is no memory. By contrast, semiotic models are based on discrete symbols and syntactic rules that have no direct relation to the laws of physics. One-dimensional strings of symbols are manipulated without regard to time or rates of change, or energy. Memory is fundamental for the existence of semiotic systems.

There is nothing wrong with trying to get as much self-organization as possible out of dynamical models, especially in the context of the origin of life before the genetic code existed. However, once coded, semiotic, description-construction exists it is not productive to minimize its significance as a heritable mechanism for harnessing dynamical laws. There is no competitive model for efficient open-ended evolution.

One current computational approach to the problem of how semiotic behavior might arise from dynamics is the study of cellular automata that can be interpreted as both a dynamical system and as a semiotic computational system (e.g., Mitchell et al. 1994). A cellular automaton is interpreted dynamically as a discrete mapping of the states of cells in a metrical space into the next state by a fixed rule that is a function of the states of neighboring cells. There are many ways to interpret the cellular automaton as a computer, but they all involve the initial state of cells interpreted as symbolic input and some later configuration of cells as the computed symbolic output. The emphasis in these models is on formal equivalences, and consequently the weakness of this approach is that there is no attempt to address how descriptions control actual physical construction, and how constructions relate to function.

The complementary approach to artificial evolution is the study of sensorimotor control in situated robots by various learning networks (e.g., Brooks 1992; Maes 1992; Hasslacher and Tilden 1995). This strategy couples the dynamics of artificial networks with the functional dynamics of sensors and activators in contact with the real physical world. Although this strategy has no direct interest in semiotic control, it is possible that such experiments may give us some clues about the origin of symbolic memory. The weakness of this approach is that this dynamic form of learning is not heritable, and consequently there is no evolvable self-replication.

16.13 When Is Downward Causation a Useful Concept?

I have argued that causation is a useful concept when it identifies controllable events or actions. Otherwise it is an empirically gratuitous linguistic form that is so universal that it results in nothing but endless philosophical controversy. The issue then is how useful is the concept of downward causation in the formation and evolution of complex systems. My conclusion would be that downward causation is useful insofar as it identifies the controllable observables of a system or suggests a new model of the system that is predictive. In what types of models are these condition met?

One extreme model is natural selection. It might be considered the most complex case of downward causation since it is unlimited in its potential temporal span and affects every structural level of the organism as well as social populations. Similarly, the concept of fitness is a holistic concept that is not generally decomposable into simpler components. Because of the open-ended complexity of natural selection we know very little about how to control evolution, and consequently in this case the concept of downward causation does not add much to the explanatory power of evolution theory.

At the other extreme are simple statistical physics models. The n-body problem and certainly collective phenomena, such as phase transitions, are cases where the behavior of individual parts can be seen as resulting from the statistical behavior of the whole, but here again the concept of downward causation does not add to the model's ability to control or explain.

A better case might be made for downward causation at the level of organism development. Here, the semiotic genetic control can be viewed as upward causation, while the dynamics of organism growth controlling the expression of the genes can be viewed as downward causation. Present models of developmental control involve many variables, and there is clearly a disagreement among experts over how much control is semiotic or genetic and how much is intrinsic dynamics.

The best understood case of an essential relation of upward and downward causation is what I have called semantic closure (e.g., Pattee 1995). It is an extension of von Neumann's logic of description and construction for open-ended evolution. Semantic closure is both physical and logical, and it is an apparently irreducible closure, which is why the origin of life is such a difficult problem. It is exhibited by the well-known genotype-phenotype mapping of description to construction that we know empirically is the way evolution works. It requires the gene to describe the sequence of parts forming enzymes, and that description, in turn, requires the enzymes to read the description.

This is understood at the logical and functional level, but looked at in detail this is not a simple process. Both the folding dynamics of the polypeptide string and specific catalytic dynamics of the enzyme are computationally intractable at the microscopic level. The folding process is crucial. It transforms a semiotic string into a highly parallel dynamic control. In its simplest logical form, the parts represented by symbols (codons) are, in part, controlling the construction of the whole (enzymes), but the whole is, in part, controlling the identification of the parts (translation) and the construction itself (protein synthesis).

Again, one still finds controversies over whether upward semiotic or downward dynamic control is more important, and which came first at the origin of life. There are extreme positions. One extreme sees the universe as a dynamics and the other extreme sees the universe as a computer. This is not only a useless argument, but it obscures the essential message. The message is that life and the evolution of complex systems is based on the semantic closure of semiotic and dynamic controls. Semiotic controls are most often perceived as discrete, local, and rate-independent. Dynamic controls are most often perceived as continuous, distributed and rate-dependent. But because there exists a necessary mapping between these complementary models it is all too easy to focus on one side or the other of the map and miss the irreducible complementarity.

16.14 Semantic Closure at the Cognitive Level

Many comparisons have been made between the language of the genes and natural language (e.g., Jakobson 1970; Pattee 1980). Typically in both genes and natural language the symbol vehicles are discrete, small in number, and fixed but structurally largely arbitrary, yet they have the potential for an unlimited number of one-dimensional expressions. These expressions are held in a memory structure that is more or less random access, i.e., not significantly restricted by time, rate, energy, and position dependence. The basic elements of the language syntax are context free and unambiguous, but as the length of expressions increases the syntax and semantics become inseparable, and when taken as a whole the semantics of the text becomes context dependent and more ambiguous, with the organism exerting more downward controls. At the many pragmatic levels the entire organism and its environment exert strong stochastic influences on meaning, function and fitness.

We know the explicit steps required to map the semiotic gene strings into the dynamics of enzyme control of rates of reactions, but almost nothing is known about the details of how the brain generates or reads the semiotic strings of natural language to produce meaning or dynamic action. Consequently, while the essential complementarity and semantic closure of semiotics and dynamics is apparent in both cases, there are certainly major differences in the structure of the memory and the dynamics and how they are coupled. First, the discrete symbols of natural language appear to be surface structures in the sense that they appear only as output of dynamic speaking or writing acts. There is no evidence that symbols exist in the brain in any local, discrete form as in the case of the gene. On the other hand, if we look at the gene symbols as input constraints on the translation and the parallel dynamic folding process as producing the output action, this is not unlike symbols acting as constraints on the input layer of a neural network and the dynamics of network relaxation as producing the output action (Pattee 1985).

16.15 Conclusion

To understand life as we know it, especially the continuous evolution of stable complex forms, it has proven essential to distinguish two complementary types of control models. One type, a semiotic model exerting upward control from a local isolated memory, and the other type, a dynamic model exerting downward control from a global network of coherent, interactive components. The semiotic model explains how control can be inherited and provides a remarkably efficient search process for discovering adaptive and emergent structures. The dynamic model suggests how the many components constructed under semiotic control can be integrated in the course of development and coordinated into emergent functions.

Neither model has much explanatory value without the other. Dynamical control models do not explain the discrete, rate-independent, orderly, heritable sequences that form the individual protein molecules, nor do semiotic control models explain how these sequences fold or self-assemble and how coordinated enzymes control the rates of specific reactions. It is true that each model alone can account for a limited level of self-organization. For example, copolymers can self-assemble more or less randomly, and by chance form autocatalytic cycles. Dynamics can also generate innumerable complex autonomous patterns. But dynamics without an open-ended heritable memory or memory without dynamic coordination have very limited emergent and survival potential. The origin of life probably requires the coupling of both self-organizing processes, but in any case, present life certainly does.

References

Anderson, P. W., & Stein, D. L. (1988). Broken symmetry, emergent properties, dissipative structures, life; Are they related? In F. E. Yates (Ed.), *Self-organizing systems—The emergence of order* (pp. 445–457). New York: Pergamon.

Bridgman, P. W. (1964). *The nature of physical theory* (p. 58). New York: Wiley.

Brooks, R. A. (1992). Artificial life and real robots. In F. J. Varela & P. Bourgine (Eds.), *Toward a practice of autonomous systems* (pp. 3–10). Cambridge, MA: MIT Press.

Bunge, M. (1959). *Causality*. Cambridge: Harvard University Press.

Conrad, M. (1990). The geometry of evolution. *BioSystems, 24*, 61–81.

Coveney, P., & Highfield, R. (1991). *The arrow of time*. New York: Ballantine.

Dummett, M. (1964). Bringing about the past. *Philosophical Review, 73*, 338–359.

Eigen, M., & Schuster, P. (1982). Stages of emerging life—Five principles of early organization. *Journal of Molecular Biology, 19*, 47–61.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Reading: Addison-Wesley.

Hart, H. L. A., & Honoré, A. M. (1958). *Causation and the law*. Oxford: Oxford University Press.

Hasslacher, B., & Tilden, M. W. (1995). Living machines. In L. Steels (ed.), *Robotic and autonomous systems: The biology and technology of intelligent autonomous agents.* Amsterdam, The Netherlands: Elsevier.

Hawking, S. (1988). *A brief history of time* (pp. 143–153). New York: Bantam Books.

Hertz, H. (1894). *Die Principien Mechanik* (English trans. D. E. Jones & J. T. Walley). *The principles of mechanics.* New York: Dover, 1956.

Holland, J. H. (1992). *Adaptation in natural and artificial systems* (2nd ed.). Cambridge, MA: MIT Press.

Jakobson, R. (1970). *Main trends of research in the social and human sciences* (pp. 437–440). Paris: Mouton/UNESCO.

Kauffman, S. (1993). *Origins of order*. New York: Oxford University Press.

Langton, C. (1989). Artificial life. In C. Langton (Ed.), *Artificial life* (pp. 1–47). Redwood City: Addison-Wesley.

Maes, P. (1992). Learning behavior networks from experience. In F. J. Varela & P. Bourgine (Eds.), *Toward a practice of autonomous systems* (pp. 48–57). Cambridge, MA: MIT Press.

Mitchell, M., Crutchfield, J. P., & Hraber, P. T. (1994). Evolving cellular automata to perform computations: Mechanisms and impediments. *Physica D, 75*, 361–391.

Nicolis, G., & Prigogine, I. (1989). *Exploring complexity*. New York: Freeman.

Pattee, H. H. (1980). Clues from molecular symbol systems. In U. Bellugi and M. Studdert-Kennedy (eds.), *Signed and spoken language: Biological constraints on linguistic form* (pp. 261–274), Dahlem Konferenzen Report 19. Weinheim: Verlag Chemie.

Pattee, H. H. (1985). Universal principles of measurement and language function in evolving systems. In J. Casti & A. Karlqvist (Eds.), *Language and life: Mathematical approaches* (pp. 268–281). Berlin: Springer.

Pattee, H. H. (1995). Evolving self-reference: Matter, symbols, and semantic closure. *Communication and Cognition—Artificial Intelligence, 12*(1–2), 9–28.

Planck, M. (1960). *A survey of physical theory* (p. 64). New York: Dover.

Schuster, P. (1994). Extended molecular evolutionary biology: Artificial life bridging the gap between chemistry and biology. *Artificial Life, 1*, 39–60.

Taylor, R. (1972). Causation. In P. Edwards (Ed.), *The encyclopedia of philosophy* (Vol. 1&2, pp. 56–66). New York: Macmillan.

Tolman, R. C. (1950). *The principles of statistical mechanics* (p. 157). Oxford: Oxford University Press.

von Neumann, J. (1955). *The mathematical foundations of quantum mechanics* (p. 351). Princeton: Princeton University Press, p. 420.

von Neumann, J. (1966). In A. Burks (Ed.), *Theory of self-reproducing automata*. Urbana: University of Illinois Press.

Weyl, H. (1949). *Philosophy of mathematics and natural science* (p. 203). Princeton: Princeton University Press.

17 The Necessity of Biosemiotics: Matter-Symbol Complementarity

Abstract Biosemiotics distinguishes life from inanimate matter by its dependence on material construction controlled by coded symbolic information. This irreducible primitive distinction between matter and symbol is necessary for open-ended evolvability and the origin of life as we know it. This type of subject/object distinction is reestablished at many levels throughout all of evolution. In physics this becomes the distinction between material laws and symbolic measurements and models; in philosophy this is the distinction between brain and mind. These are all emergent epistemic distinctions, not ontological dualisms. The origin of life requires understanding the origin of this symbolic control and how inanimate molecules become functional messages. I discuss the necessary physical conditions that would allow such evolvable symbolic control of matter to arise.

Keywords Self-replication • von Neumann • Symbolic control • Semantic information • Measurement • Constraints • Epistemic cut • Protein folding

17.1 Life Depends on Semiotic Controls

We easily agree with Einstein that a Beethoven symphony cannot be appreciated as only "a graph of air pressures," although in principle it has such a physical description. In the same way we understand Bohr that, "You don't explain a tea party by quantum mechanics." On the other hand, it is not so easy to understand why you cannot adequately explain genetics with biochemistry or enzyme catalysis with quantum mechanics. Because we believe no events at tea parties, in genes, or in enzymes violate any physical laws we might assume that their descriptions differ

---

Reprinted from *Introduction to Biosemiotics*, Marcello Barbieri, Ed. Dordrecht, The Netherlands: Springer, 2007, pp. 115–132.

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7,
DOI 10.1007/978-94-007-5161-3_18, © Springer Science+Business Media Dordrecht 2012

only in their degrees of complexity. What biosemiotics illustrates is that symbolic controls are categorically different from laws and that they are irreducible to physical laws even though their material vehicles obey the laws and have a correct physical description.

What we need to understand is that physical laws are universal and must apply to all conceivable systems. Therefore laws are empirically moot with respect to any particular system until its particular initial conditions are specified. This requires information, and physical laws cannot specify this information. In physics jargon symbol systems are special types of initial condition called boundary conditions or constraints (Polanyi 1968; Pattee 1972). Consequently an adequate explanation of any living organism requires more than a detailed lawful physical description or merely the confirmation that the laws of nature are always inerrantly followed. One must explain how informational constraint structures locally control the universal physical laws so as to propagate and evolve.

All living organisms exist by virtue of hierarchies of control by informational constraints. This is the case at all levels, from the genes, to development, to sensorimotor controls, to abstract thinking, and to our technical artifacts. Symbol systems are rate-independent informational constraints that control rate-dependent dynamics by means of coding systems.

To understand what this implies one must first recognize that physical laws are universal and objective. This means that the fundamental principled requirement for a law of nature is that it is as independent as possible of all conceivable individual organisms and observers. Consequently, physical laws are based on invariance and symmetry principles that guarantee the irrelevance and impotence of any observer, organism, or mechanism to affect the laws. In other words, *physics focuses on all those universal regularities of nature over which organisms and observers have no control.* Physical laws are universal and inexorable. By contrast, the study of *biology focuses on those specific events over which the organisms and observers have local control.* Beginning with the organism's specific catalytic rate control by enzymes, evolution progresses by elaborating and testing many types of controls at many hierarchical levels. Over the course of evolution organisms have gradually increased their ability to control their internal and external environments on which their survival depends.

Survival is the ultimate function of controls, but unfortunately controls do not assure survival. In the case of humans the brain through the freedom of language and the prostheses of technology has developed controls with such Promethean powers that the art of artificial control may turn out to have exceeded what is adaptive as a survival strategy for our species.

17.2 The Relation Between Physical Laws and Control Constraints

There is a clearly problem of language here that creates confusion. What does it mean to say that universal inexorable physical laws over which organisms can have no control are in fact controlled by individual organisms? The answer requires understanding a distinction fundamental to all physical theory, the primitive separation of the laws themselves that are universal and inexorable, and initial conditions and constraints that are local and controllable, and that must be measured to have any effect. Eugene Wigner calls this principled distinction between laws and initial conditions "Newton's greatest discovery." That is because it is an epistemic necessity that is essential for all conceivable physical laws, not just Newton's laws.

Briefly, the idea is this. The universe and all systems within it are assumed to run according to universal laws whether or not observers or life exist. The mathematical descriptions of these laws are interpreted by ontological concepts of space, time, matter and energy but the laws themselves do not include the epistemological concepts of measurement and control events. However, measurement is essential if we want to predict any consequence of laws on a specific observable system. There must be measurement of initial conditions and the measurement process requires local control constraints of a measuring device or instrument.

Physical laws and initial conditions are therefore necessary irreducibly complementary categories. That is, neither can be reduced to, or derived from, the other. Measurement instruments and control constraints are special, usually complex, structures with initial conditions that are largely arbitrary. In practice measurements and controls are distinguished from the action of physical laws by how time and energy enter their descriptions. Fundamental physical laws are time and energy dependent in a mathematically rigorous sense. That is, the equations describing these laws require the concept of time-derivatives or rates of change of the states and energies of the system[1]. Also, the fundamental microscopic laws are time reversible. This physical time, sometimes called "real time," and the rates described by time derivatives are intrinsic to natural laws and are not controllable, although they may be different when measured by different observers in relative motion.

The concept of *control* of rates does not apply to universal laws but only to local structural constraints. The classical example of both rate control and time measurement is a clock. By contrast to the real-time of laws, clock-time depends on some form of local structure or constraint. We speak of clocks *measuring* time intervals but, unlike laws, clocks do not have an intrinsic rate independent of how we measure it. Also, unlike microscopic laws, measurement and control are irreversible concepts. Clocks function only by measuring local periodic structures such as a pendulum with an escapement or counter. Of course the pendulum swings according to laws, but its period depends on its length, and that is entirely arbitrary boundary condition. Escapements, whether mechanical, electronic, or chemical, can be said to control the rate at which energy "escapes" or is dissipated from the driving source, and these constraints are also arbitrary[2]. Some form of measurement is a necessary component of any functional control process.

The point I want to emphasize here is that we say a clock is a control constraint only by virtue of its locally "escaping" the inexorable time, rate, and energy dynamics of physical laws. In other words, the laws exist in time but cannot make measurements of time. Within wide limits imposed by natural laws, a clock keeps its own arbitrary time and runs at its own arbitrary rate. *This concept of local "escape" is important because life depends on it.* Enzymes control the rates of specific chemical dynamics in all of life allowing local organisms to locally escape the universal rates we associate with unconstrained physical laws. The existence of an isolated catalyst that controls rates of reaction is not considered as functional. Function, as I use the term, applies only coordinated controls initiated by organisms or executed by their artifacts.

17.3 Rate-Independent Constraints; Symbol Systems

Biosemiotics recognizes many levels of control. Rate control, as in enzyme catalysis, is only the first level of control constraints. Symbolic constraints are a type of record that requires higher levels of organization. Ernst Mayr (1982) has often pointed out that biology is not explained by natural laws because life requires the concept of an adaptive evolutionary *history*, an actual record in the organism that while obeying natural laws cannot be derived from these laws. Records are a special form of constraint that can “instruct” controls. Polanyi (1968) classifies these structures in physical terms as special types of boundary conditions that he aptly describes as “harnessing” the laws.

The word “history” has two profoundly different usages. The looser meaning is simply an implied ontological sequence of events, as in the history of the universe, or geological history, where there is no explicit record other than the actual events or structures themselves. The more specific meaning of history as used by Mayr requires a separate record of events. This latter specific meaning that is essential for evolution implies an epistemic record that is a representation or description distinct from the events that it records. In all known living systems, the genes are such historical records of innumerable adaptive natural selection processes. The relative simplicity of the record itself (the DNA) is deceptive. What is important is that for a record to have any function or meaning requires complex coding, reading and interpreting mechanisms. Along with measurement and control the concepts of biological information and instruction are not a part of physical theory in so far as they are functional concepts. That is, we cannot identify a molecule as informational unless we can identify how it is interpreted by the organism and how it functions in the organism[3]. The question remains, how does symbolic information actually get control of physical systems when it appears to be a separate category?

17.4 Physical Laws Cannot Address This Question

This matter-symbol separation has been called the epistemic cut (e.g., Pauli 1994). This is simply another statement of Newton’s categorical separation of laws and initial conditions. Why is this fundamental in physics? As I stated earlier, the laws are universal and do not depend on the state of the observer (symmetry principles) while the initial conditions apply to the state of a particular system and the state of the observer that measures them. What does calling the matter-symbol problem “epistemological” do for us? Epistemology by its very meaning presupposes a separation of the world into the knower and the known or the controller and the controlled. That is, if we can speak of knowledge *about* something, then the knowledge representation, the knowledge vehicle, cannot be in the same category of what it is about.

The dynamics of physical laws do not allow alternative paths between states and therefore the concept of information, which is defined by the number of alternative states, does not apply to the laws themselves. A measurement, in contrast, is an act of acquiring information about the state of a specific system. Two other explicit distinctions are that the microscopic laws are universal and reversible (time-symmetric) while measurement is local and irreversible. There is still no question that the measuring device must obey the laws. Nevertheless, *the results* of measurement, the timeless semantic information, cannot be usefully described by these time-dependent reversible laws (e.g., von Neumann 1955).

17.5 The Epistemic Process in Biology

The problem is that physics greatly simplifies the matter/symbol relation by more or less arbitrarily making an epistemic cut. This avoids as far as possible the influence of measurement on the state of the measured system. Whenever an attempt is made to include the measuring device in the system it becomes the notorious "measurement problem" on which there is still no consensus.

The origin of life must address the question: How did this separation, this epistemic cut, originate? As Hoffmeyer (2000) has pointed out, the assumed sharp conceptual epistemic cut between these highly evolved categories of knower and known makes it difficult to imagine how life began and how these two categories separated at primitive levels. The epistemic cut is often treated as a conceptual discontinuity. Indeed it is difficult to imagine a "gradual cut." How does a reversible dynamics gradually become an irreversible thermodynamics? How does the concept of counting discrete units transform into the concept of a continuum (Zeno's paradox of motion)? How does a paradigm shift from classical determinism to quantum indeterminism occur gradually?

The problem arises acutely with the genetic code. A partial code does not work, and a simple code that continuously works as it evolves is hard to imagine. In fact, this is a universal problem in evolution and even in creative thought. How does a complex functioning set of constraints originate when no subset of the constraints appears to maintain the function? At least in the case of thought we can trace some of the history, but in the origin of life we have no adequate history. Even in the case of creative thought, so much goes on in the subconscious mind that the historical trace has large gaps.

The problem is that conceptually the epistemic cut divides the world in two, and the central problem is how the two worlds are connected. As C. S. Peirce has emphasized, all symbol systems are necessarily triadic systems, and the epistemic cut itself is actually a complex process. It corresponds to the *interpretation* that relates the symbol to its referent. In the cell this is an enormously complex process of transcription, translation, synthesis, folding, distribution, and selective control of many proteins. How this coordinated interpreting system originated is the central problem of the origin of life.

17.6 Symbolic Control is Necessary for Evolvability

The categorical distinctions between matter described by physics and matter functioning as symbol are different at each level of biological organization. The distinction needs to be made clearly at each hierarchical level or conceptual and terminological confusion will result. It is easy to distinguish symbols at highly evolved levels like symphonies and tea parties. The words on this page are clearly symbols. Their material embodiment is arbitrary. The font is not relevant, nor do we pay attention to their material embodiment, whether they are displayed on a liquid crystal screen, a cathode ray tube, or ink on paper. Even the language we are using is arbitrary.

It is not so easy to see that the DNA of genes is also an arbitrary embodiment of a record because it happens to be the only one we know from life on earth. However, within the fields of exobiology and artificial life studies the arbitrariness of DNA is generally assumed. Many other copolymer strings or even bit strings in a computer could be interpreted or translated by a suitable coding mechanism to synthesize the same proteins as a DNA sequence.

Why is this arbitrariness of symbols essential for open-ended evolution? The most obvious property of highly evolved symbol systems such as natural language and mathematics is their enormous open-ended variety that is not limited in any significant way by physical laws. This independence is also illustrated by the fact that, unlike physical laws, the function and meaning of symbols is not dependent on the rate at which they are written or read. A mathematical proof does not depend on how long it took to produce or to read. The same it true of a work of literature. In other words, the basic observables of physical laws, space, time, matter, energy, and rates of change, have no significance for the semantic information of symbol systems. The symbolic expressions of physical laws are "about the laws" but the mathematical symbols that describe the laws do not appear to be restricted by the laws. It is just this arbitrariness that allows organisms freedom to harness laws. The necessity of symbols for open-ended evolution was first discussed by von Neumann (1966) in his lecture on the logic of self-replication.

17.7 Von Neumann's Description and Construction

Von Neumann was the first to argue that the two categories, *symbolic description* and *material construction*, are essential for self-replication that is capable of open-ended evolution. His argument was entirely abstract and by no means logically complete. It explicitly abstracted away matter, energy and all physical laws. I will first elaborate on von Neumann's logic and then.

I will take up the necessary *physical* conditions to realize this logic, or what he thought, "may be the more important half" of the problem.

Von Neumann's logical argument for the necessity of symbols as distinct from dynamics in self-replication was informal and largely intuitive. Nevertheless, if you understand his argument you will find it hard to think how evolvable self-replicating units could work any other way. The motivation for his argument was to understand the "threshold of complication" that would allow systems to evolve increasing complexity rather than wearing out or decaying. His logic is all the more remarkable because it correctly predicted how cells actually replicate before the discovery of the mechanisms of genetic description, coding and protein synthesis. Von Neumann began by observing that the medium of communication that feeds a material automaton is completely different than the automaton itself or its output. This was his recognition that symbols are a different category than matter. He also recognized that this was important for general-purpose computers, what is called the software-hardware distinction.

Von Neumann emphasized the "completely decisive property of complexity, that there exists a critical size below which the process of synthesis is degenerative, but above which the phenomenon of synthesis, if properly arranged, can become explosive." He was thinking of biological evolution and its open-ended variety. The essential condition here is that the individual self that is being replicated must be only one of an indefinite number of different potential selves all of which can be replicated by the same process. This raised two questions: (1) what defines the set of all possible individual selves that potentially can be replicated? And (2) how do you describe or represent the individual *self* that is being replicated?

Logic will get you only so far with these questions. For example, the concept of replication implies assembling or constructing a new individual that is like another. Von Neumann realized that how this construction can be done will depend on the nature of the available parts and on how the parts are to be assembled. He saw that if the parts were too elementary, like atoms, then both the description and construction would be a long and complicated process, while if the parts were too complex, like robots or rabbits, then there would be no real problem. He called this the "parts problem" and abstracted away the matter and energy of real construction by defining some functional operations on parts, like recognizing, moving, cutting, joining, etc., that are to be symbolically represented. There is a great amount of arbitrariness in these choices of parts and operations, but as we shall see, the basic logical separation of symbolic description and material construction does not depend on these choices.

The more fundamental question is how you make sure the replicated individual is like the original. How do you construct a copy of an organized structure made up of parts from a reservoir of these parts? There are two approaches. One is to identify the original parts themselves by *inspection* and then assemble the corresponding parts to form the copy. The other approach is to use a *description* of the original that when interpreted amounts to instructions enabling the assembly of the parts in the copy. Note that the concepts of inspection and description require an epistemic cut that separates the object being inspected or described and the record of the inspection or description. Both of these methods have advantages and disadvantages that go beyond logic and depend on the physical nature of space, time, and the nature of the parts. Von Neumann using heuristic reasoning found that taking advantage of both approaches gives the most promising results, and in fact we now know that both approaches are used in all living systems in the way that von Neumann proposed.

17.8 Von Neumann's Logic of Self-Replication

Following these intuitions, von Neumann began simply by postulating the existence of both symbolic and material components in the forms of a *description* and a *constructor*. The constructor would both interpret and construct what was described using parts from a reservoir. The constructor was universal with respect to an open-ended set of descriptions one of which he assumed could be the description of the constructor itself. In his notation, A was the material constructor and $\phi(A)$ was the description of the constructor. If the description $\phi(A)$ was fed to the constructor A, then A would construct a copy of itself, A′. We can symbolize this as $\phi(A) \rightarrow A = A'$. This is not self-replication because the description $\phi(A)$ has not been replicated. One might at first think that to copy the description we would simply feed the constructor a description of the description, $\phi(\phi(A))$, but this leads to an infinite regress since that description must also be copied, and so on.

This leads to the crucial recognition that *a symbolic description, whatever form it may take, has a physical structure that is independent of its interpretation*. In other words, to *read* the description means to *interpret* the description. To *copy* the description means *not to interpret* the description but to copy only its physical structure. Since the description is quiescent, copying can be done by inspection or by some template process. The constructor is defined to only interpret the description, so it is necessary to add another component, B, called the copier and its description $\phi(B)$. We then can write $\varphi(A+B) \rightarrow (A+B) = \varphi(A'+B') \rightarrow (A'+B')$. This is almost self-replication except it is ambiguous. What is missing is how the new descriptions and constructions are related. Von Neumann "solved" this logically by creating a new control component, C, that takes care of housekeeping details such as inserting the new description into the new hardware constructor and separating the offspring from the parent. This component, C, amounts to what is called the operating system of a computer that takes care of the software-hardware relationship.

Von Neumann's logic and computer analogies are by no means a clear solution to the material semantics of cells. In the cell we know that the control required for cell division is a very complex process that is not yet fully understood. But the essential evolutionary consequence of von Neumann's logic is that now any additional description, D, of some new structure or function when added to this basic description will be constructed and incorporated into all future generations:

$$\varphi(A+B+C+D) \rightarrow (A+B+C) = \varphi(A'+B'+C'+D') \rightarrow (A'+B'+C'+D')$$

This is as far as von Neumann's logic takes us. The main point of his logic is that open-ended evolution requires more than a complex time-dependent dynamics and complex chemical reactions. There must be a time-independent passive memory that by means of a coded description controls the dynamical rates of specific constructions or chemical syntheses. What I will now take up are the physical requirements that would allow such a complicated symbol-matter logical scheme to actually work in a reasonably effective way. I repeat that I am not solving the origin problem. Von Neumann himself had no clue. He thought, "That such complex aggregations should occur in the world at all is a mystery of the first magnitude." In my view, the place to look for clues is in the actual physical requirements of symbol systems where we may imagine simpler systems than we find in today's highly evolved organisms that satisfy these requirements.

17.9 Von Neumann's "More Important" Question

Von Neumann was fully aware that logic alone was not adequate to explain cells. He warned: "By axiomatizing automata in this manner one has thrown half the problem out the window and it may be the more important half. One does not ask the most intriguing, exciting and important questions of why the molecules or aggregates that in nature really occur … are the sorts of thing they are, why they are essentially very large molecules in some cases, but large aggregations in other cases."

Von Neumann's use of *inspection* and *description* are really generalizations of highly evolved cognitive activities that need to be more precisely defined in the context of the simplest replicating unit. Copying by inspection means using physical interaction with the object directly without the use of symbols, codes, translation, or interpretation. Casting from a mold and template matching are such direct processes as in base pairing in copying nucleic acids and the binding of a substrate by an enzyme. I should emphasize here that the physical interaction of base pairing and substrate binding are not in themselves functional or semiotic processes. It is only by virtue of their roles in the overall process of self-replication that they are interpreted as functional. Such material matchings might be interpreted in Peirce's terms as iconic signs.

A description, on the other hand, requires more complicated physical interactions that couple the description to what it stands for, its referent. This interaction in the context of self-replication can be called a code or an interpretation, and because the code constraints are themselves constructed from a description they are not determined by physical necessity. It is implicit in the concept of a code that it must apply to more than one description. In fact, to allow evolution the code must apply to an open set of potential descriptions. Again I emphasize that only by virtue of its potential function for an individual's survival can this be distinguished as a semiotic process. This chemical arbitrariness in the coding enzymes Jaques Monod (1971) calls the "principle of gratuity." It is also this construction from a description that Barbieri (2004) calls "artifact-making," a distinguishing characteristic of life. It is because of this freedom or lack of physical necessity that genetic symbol systems and the novelties of evolution have no adequate physical explanation even though they can in principle be correctly described by physical laws in every detail.

It is not clear that von Neumann saw this point since he was concerned with the logic, not the physics. However, he did argue that a description had the advantage of being quiescent, relatively time-independent, and free of the dynamics of the system it describes. It could then be copied by direct inspection. On the other hand, copying a dynamic system by direct inspection in real time would run into a problem with the parts continually changing in time. How would the system choose what state should be copied in that case? He also suggested that a complete and detailed inspection, including inspecting the inspection components themselves, would probably lead to logical antinomies of self-reference. He did not elaborate on this, but he may have been thinking of the measurement process in physics where he showed elsewhere that measuring the initial conditions of the measuring device itself leads to an infinite regress. Only by choosing at some point to make the distinction between the system being measured and the measuring device, i.e., an epistemic cut, can this regress be terminated (von Neumann 1955).

17.10 Physical Requirements for Efficient Memory

The physical conditions necessary for memory storage are relatively simple to state as contrasted to the conditions for writing and reading of memory. The first condition is that there exist many inherently equiprobable constraint structures with adequate stability. Equiprobable means that the structures are energy degenerate or the energy of each state is the same. These states need not be exactly the same energy as long as the energy differences do not significantly affect the setting of the state by writing or the communication of the state by reading. One-dimensional copolymers and linear symbol strings are the simplest common physical structures satisfying these conditions. Such relatively time-independent memory structures function as long-term, high capacity storage.

Memory structures can also exist physically in one, two, three dimensions, or in n-dimensional networks but explicit syntax for access must be supplied. The advantages of the linear sequence memory, like nucleic acids and Turing machine tapes, and language text are (1) open-endedness or extendable capacity, (2) uniformity and simplicity of writing and reading, including ease of random access, (3) universal coding for all sequences, (4) relative isolation from the dynamics that it controls because of coding or the interpretation process. In the context of the origin of life, copolymer chains are the simplest abiogenic structures that have the necessary stability and potential memory capacity. The disadvantages of linearity are (1) lack of parallel processing or associative access, (2) low density of information storage, and (3) the necessity for an explicit code to couple one-dimensional energy degenerate sequences to the energy-dependent three-dimensional dynamics.

One can also define analog memory and codes as in analog computation. Analogs need not involve discrete symbols. This has been suggested by Hoffmeyer and Emmeche (1991), Juarrero (1998), Hoffmeyer (1998) and Barbieri (2003) in contrast to discrete or digital memories and codes. The problem with analogs is that they are all special purpose structures like individual molecular messengers that have limited informational capacity and that have no common code or interpreting process, as do genetic sequences. An autocatalytic or metabolic network may be interpreted as containing an implicit informational dynamics, but lacking an explicit passive memory structure and code it is difficult to imagine any open-ended evolvability. On the other hand, as Hoffmeyer (2000) suggests, some form of implicit analog codes may have existed as precursors of the explicit discrete codes of present life.

17.11 Physical Requirements for Coding and Construction

In even the simplest existing cells the steps from the symbolic base sequence in DNA to a functioning enzyme are too complex to have originated without simpler intermediate stages. However, to control construction or synthesis, even the simplest one-dimensional discrete-state memory storage that exists by virtue degenerate energy states, must somehow control the rates of specific dynamical interactions. This means that the linear degeneracy must be broken. This must be done by new interactions between the linear storage elements. In present cells this is a complex process that requires several steps. First, the DNA sequence is transcribed to messenger RNA by template copying. Next the coding enzymes and transfer RNAs translate the base triplet code to the corresponding amino acids that are then joined in sequence by the messenger RNA and ribosome machinery. Finally, the one-dimensional sequence folds into a functioning enzyme. In this process there are cases of descriptions and constructions by both template inspection and coded descriptive translations.

The discovery of enzymatic RNA made it possible to imagine a much simpler translation process in which RNA can function both as a constructing enzyme and as a symbolic description of an enzyme. By description I mean a passive structure that can be copied by template inspection, and by construction I mean a dynamic catalytic process that joins molecules by strong, covalent bonds. The main point is that this double function is only possible by virtue of the two configurations of RNA, the passive one-dimensional sequence memory and the folded three-dimensional active ribozyme.

17.12 The Physical Requirements for Folding and Function

Folding transformations are the most fundamental semiotic processes in all living systems. Folding is fundamental because it is the process that transforms the passive symbolic gene sequences into the dynamic rate-control of enzymes. Folding transforms what are essentially rate-independent syntactically coded sequences into rate-dependent functional controls. Protein folding is a highly parallel process with so many degrees of freedom that is difficult to model even on supercomputers. Physically to describe folding in any structure requires two types of bonds, strong bonds that preserve the passive topological structure of what is folded, and weaker bonds that acting together hold the active folded structure in place.

This physical requirement follows from the logical definition of "folding." For example, to fold a sheet of paper means forming a three-dimensional shape without changing the two-dimensional topology of the sheet by tearing or gluing. As long as the strong-bond topological sequence structure is energy degenerate it can serves as an informational constraint or a passive memory. Folding removes this degeneracy by allowing new weak bond interactions between the elements resulting in an active enzyme. A *physical description* of protein folding is an energy minimization process or a relaxation of many weak bond interactions under the constraints of the strong bonds holding the linear sequence together (e.g., Frauenfelder and Wolynes 1994).

How should we describe the semiotics of this process? I want to distinguish the physics and the semiotics. First, I defined a condition for symbolic information storage as a physically indeterminate (energy degenerate) structure. I assumed that all symbol vehicles obey physical laws and have, in principle, a physical description, but as I explained that does not imply that symbol structures are physically determined. Quite the contrary is the case. Such a degenerate sequence structure can have an immense number of physically indeterminate sequences. Therefore the interpretation or function of any such semiotic or informational sequence is literally metaphysical (beyond physics).

The actual folding process, on the other hand, is an entirely physical process of minimizing the energy under the semiotic constraints of the sequence. In other words, the strong-bonded sequence can be called informational because it is one of many physically equivalent alternative sequences, while the folding dynamics itself is not informational because no new information is added in the process of minimizing the energy. (There are special cases where folding information may be added from scaffolding molecules.)

17.13 The Semiotic Closure Requirement for "Self"

How do we define the individual system that is interpreting the information? We need an objective criterion for what "self" is doing the interpreting and replicating, because there are innumerable energy degenerate structures that are not descriptions and many catalytic events that are not functional. What additional conditions are required to satisfy a *physical* implementation of the *logical* "self" that reads and interprets descriptions and constructs and assembles parts in von Neumann's formal self-replication.

The essential logical requirement for self-replication that von Neumann described is that all the components that implement description, translation, and construction are themselves described, translated and constructed within the “self” that is being replicated. This amounts to a *logical closure* that defines a “self.” Physically this requires elaboration. There is more to the strong and weak bond requirement than the ability of the weak bonds to cause the strong bonds to fold into a functioning enzyme. The strong bonds also stabilize the passive memory and the integrity of the primary structure of enzymes. The weak bonds bind the enzyme to its substrate and control the rate of catalyzed strong bond formation. In effect, the strong bonds form the skeleton for both descriptive and constructive molecules while the coordinated organization of weak bonds define the shapes necessary to control the strong bonds, both the strong bond folding and individual strong bond formation or breaking.

These are the *physical* conditions required to implement von Neumann’s logical closure. I have called this *semantic closure*, but Luis Rocha (2001) has more accurately called it *semiotic closure* because its realization also includes the syntax and pragmatic physical control processes. This complex interrelationship of strong and weak bonds is the minimum physical requirement that allows the realization of von Neumann’s quiescent symbolic description and dynamic material construction. Of course the actual physical forces come in more than two strengths and evolution has refined structures at many hierarchical levels using different types of forces. Many types of strong and weak bonds enter into the complex process of folding (e.g., Wolynes, et al. 1995).

17.14 Evolution Requires Population Distributions

Based on the concept of semiotic closure, I would define an interpreter as a semiotically closed localized (bounded) system that survives or self-reproduces in an open environment by virtue of its memory-stored constructions and controls. That distinguishes interpreters from inanimate physical systems that evolve dynamically simply because they follow the memoryless state-determined laws of nature. I believe that this minimal concept of interpreter is consistent with Ghiselin’s (1997) more elaborate definition of an “individual” that also applies to higher levels, like species. However, just as there are no single symbols that have meaning, so there are no single interpreters capable of efficient evolution.

Symbols exist only in the context of codes and interpreters. Symbols are recognized in an individual interpreting system just because they function in propagating the system. But we cannot stop there. We immediately see that “propagating a system” is ambiguous. The individual interpreter is not enough. The whole idea of evolution by variation and natural selection depends on a *population* of individuals that can differ in their heritable memories. This leads directly to the central issue of evolution: what kinds of symbolic descriptions, control constraints and material constructions promote survival of *populations*? Of course there is no predictable answer to this question except the course of evolution itself. All we can do is look carefully at what is actually going on in existing organisms, and see if we can discover some answers to von Neumann's question of why the molecules are the sorts of thing they are. I will mention some properties of memory, codes, symbolic control, and material construction that studies suggest promote efficient evolutionary search and natural selection.

17.15 Requirements for Efficient Search and Selection

After asking this question von Neumann remarked that it was "a very peculiar range" for the parts since they were many orders of magnitude larger than the physically elementary particles. He did not discuss this except to suggest that the size had to do with the reliability of control since in automata there is a direct correlation between number and size of parts and reliability. A certain level of reliability is certainly one requirement in order to prevent error catastrophe, but another way to look at the question is in terms of function. How small could an enzyme be and accurately bind a substrate and catalyze a specific single bond. It would have to be a large enough structure to establish a shape with the necessary specificity to recognize a substrate by folding up a linear chain. Simple models suggest that of the order of 100 amino acids is necessary.

This size creates two fundamental problems. The first problem is that the number of copolymer sequences of such lengths is immense, well beyond actual enumeration. One of the oldest, non-religious arguments against Darwinian evolution is the apparent improbability of chance mutations producing any successful protein, let alone a species. This is still an argument used by "intelligent design" advocates. This argument is based on the assumption of the sparseness of functional sequences and the immensity of the search space. The weakness of this argument is that the actual probabilities of the events in question are largely unknown.

Formulated in biosemiotic terms, to address this problem we need to know what fraction of the innumerable potential symbol strings in a genetic memory has some meaning or function when expressed by a population of individual interpreters. We need to know how the enormous space of sequences maps into the space of biological functions. The second classical problem is that functions appear to be discretely separated. That is, one function does not smoothly transform into another function. This results in the so-called trapping problem on a function or fitness landscape.

Both these problems have been studied extensively, greatly assisted by the use of computational models. Of course, there are no pure theoretical answers. Some basic empirical knowledge is required of the actual polymers that form the memory sequence space, the nature of codes that map to protein sequences, the nature of folding, and the nature of the constructive or controlling enzymes. The auspicious discovery of molecular genetics was that many mutations are neutral with respect to function and fitness (Kimura 1983). Along with the redundancy in the genetic code, this neutrality permits searches over a wide region near a function optimum or a local fitness peak thereby alleviating the trapping problem. Trapping is also greatly reduced by the large number of saddle regions that increases with the dimensionality of the memory sequence space (e.g., Kanerva 1988). This lends weight to the concept of quasispecies and the advantages of mutation rates near the error threshold (Eigen 1971; Eigen and Schuster 1979).

This search problem has been studied extensively for the simple RNA worlds of sequences and their folding (e.g., Schuster, et al. 1994; Schuster 1998; Crutchfield and Schuster 2003). Again the mapping of passive memory sequences to shapes that could function as enzymes appears to be highly redundant with many sequences resulting the same three-dimensional shape. Furthermore, these sequences are distributed more or less uniformly over the entire sequence space. This means that a random search need not find just one needle in a haystack, but only one of many needles uniformly distributed over the whole haystack. That is, wherever a random search begins in sequence space, it appears likely that a description of a useful molecule will be found nearby.

17.16 Analogies and Disanalogies of Genetics with Natural Language

Biosemiotics is the study of all forms of signification and communication. It recognizes that life is distinguished from the nonliving world by its dependence on signs and symbols. However, of the innumerable examples of pattern recognition, recording, signaling, and communication throughout all levels of living organizations only two clear examples of open-ended, creative language systems exist, the genetic language and natural languages. The similarities of genetic sequences and natural language have struck linguists as well as biologists and physicists (e.g., Jakobson 1970). These two languages can be characterized by (1) a small, fixed alphabet, (2) one-dimensional expressions in discrete sequences, (3) an immense sequences space with no significant restriction or bias from physical laws, (4) expressions not limited in what they can potentially describe by what currently exists, (5) the interpretation of sequences, their function or meaning is complex requiring highly parallel processing. In the case of genetic sequences, the essential step is folding in which many strong constraints and weaker forces act in parallel. In the case of the brain, millions of neurons are involved in interpreting even the simplest expressions (Pattee 1980).

Natural language structure also illustrates the strong and weak bond principle, not with a hierarchy of physical forces but with a hierarchy of rules. The lexical rules are the most rigid beginning with the alphabet and the words in the lexicon. The grammar rules are weaker than the lexical rules in the sense that syntax cannot control or modify the alphabet or the dictionary. The semantics of the text does not generally alter syntax. We usually assume our writing will not change the basic meanings of words or the grammar rules depending on what we write. Similarly the sequence or meaning of the code's base triplets is not changed by the functions of enzymes they describe. Notably however, both languages have evolved exceptions to these rules, the genetic system with special editing enzymes, reverse transcription and cutting and splicing, and natural language with freedom to invent metaphors, add new words, and to violate grammar rules with figures of speech.

Of course there are enormous differences between these languages both in their embodiments, their stability, and in their range of meanings which one would certainly expect considering they originated only at the very beginning and the end of the evolutionary time scale. The genetic language began with the origin of life, and it took 4 billion years of evolution to create brains with the capability to create natural languages. The genetic language can be called highly successful in creating adaptive functions that have kept life going over this enormous time span.

As I suggested at the beginning of this paper, whether language will turn out to be a long-term evolutionary success is not at all obvious. We often refer to natural language as the defining characteristic of human intelligence. The power of language has dominated history and shaped all our cultures. Human language has not existed for more than 100,000 years and it is quite possible that it has become too persuasive for generating myths and wishful thinking that avoid basic survival necessities for the species. Also, the technology that depends on language now allows us to design genetic messages that satisfy immediate human desires rather than long-term survival of the species. Assuming humans survive the dangers of natural language and technology, one wonders what higher level of languages might evolve in 100,000 years. If humans do not survive natural language and technology, one wonders what alternative biosemiotic structures might evolve in its place.

Notes

Sections of this paper are edited and updated selections from H. H. Pattee, The physics and metaphysics of biosemiotics, *Journal of Biosemiotics* 1(1), 223–238 (2005).

1. This statement applies to the relatively narrow range of time and energy domains within which living organisms have been found to exist on earth. Fundamental particle and cosmological theories are far outside these domains, although the possible relevance of these theories to other conceivable forms of life is an open question.
2. Natural periodic motions like the rotation of the earth and the emission frequencies of atoms also serve as a reference for clocks, but without arbitrary and often elaborate dissipative constraints the function of any clock, that is, the measurement of time, does not occur. The word control is also sometimes used in a more general sense to describe parameters in physical systems where no function or measurement is involved.
3. Physicists and engineers often use information in a structural rather than functional sense because of its formal relation to the entropy of a system. *Structural information* is defined in communication theory (e.g., Shannon and Weaver 1949). Also in quantum processes one may think of structural information being transferred from the quantum system to the observing system (e.g., Zurek 1990). I am restricting my usage to *semantic information* that functions in the survival of biological organisms and populations.

References

Barbieri, M. (2003). *The organic codes. An introduction to semantic biology*. Cambridge: Cambridge University Press.

Barbieri, M. (2004). Life is "artifact-making". *Journal of Biosemiotics, 1*(1), 81–101.

Crutchfield, J. P., & Schuster, P. (Eds.). (2003). *Evolutionary dynamics: Exploring the interplay of selection, accident, neutrality, and function*. New York: Oxford University Press.

Eigen, M. (1971). Self organization of matter and the evolution of biological macromolecules. *Naturwissenschaften, 58*, 465–523.

Eigen, M., & Schuster, P. (1979). *The hypercycle – a principle of natural self-organization*. Berlin: Springer.

Frauenfelder, H., & Wolynes, P. G. (1994). Biomolecules: Where the physics of complexity and simplicity meet. *Physics Today, 47*, 58–64.

Ghiselin, M. T. (1997). *Metaphysics and the origin of species*. Albany: University of New York Press.

Hoffmeyer, J. (1998). The unfolding semiosphere. In G. Vijver, S. Salthe, & M. Delpos (Eds.), *Biological and epistemological perspectives on selection and self-organization*. Dordrecht: Kluwer.

Hoffmeyer, J. (2000). Code-duality and the epistemic cut. In C. Jerry & G. Van de Vijver (Eds.), *Closure. Emergent organizations and their dynamics* (pp. 175–186). New York: Annals of the New York Academy of Sciences.

Hoffmeyer, J., & Emmeche, C. (1991). Code duality and the semiotics of nature. In M. Anderson & F. Merrell (Eds.), *On semiotic modeling*. New York: Mouton de Gruyter.

Jakobson, R. (1970). Linguistics. In *Main trends of research in the social and human sciences I* (p. 438). Paris/The Hague: Mouton-UNESCO.

Juarrero, A. (1998). Causality as constraint. In G. Vijver, S. Salthe, & M. Delpos (Eds.), *Evolutionary systems. Biological and epistemological perspectives on selection and self-organization*. Dordrecht: Kluwer.

Kanerva, P. (1988). *Sparse distributed memory*. Cambridge: MIT Press.

Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge: Cambridge University Press.

Mayr, E. (1982). *The growth of biological thought*. Cambridge: The Belknap Press of Harvard University Press.

Monod, J. (1971). *Chance and necessity*. New York: Knopf.

Pattee, H. H. (1972). Laws and constraints, symbols and languages. In C. H. Waddington (Ed.), *Toward a theoretical biology 4. Essays* (pp. 248–258). Edinburgh: Edinburgh University Press.

Pattee, H. H. (1980). Clues from molecular symbol systems. In U. Bellugi & M. Studdert-Kennedy (Eds.) *Signed and spoken languages: Biological constraints on linguistic forms* (pp. 261–273). Dahlem Konferenzen, Berlin: Verlag Chemie GmbH.

Pauli, W. (1994). The philosophical significance of the idea of complementarity. In C. P. Enz & K. von Meyenn (Eds.), *Writings on physics and philosophy* (pp. 35–48). Berlin: Springer, (see p. 41). First published under the title "Die philosophische Bedeutung der Idee der Komplementarität" in *Experientia* 6(Heft 2), pp. 72–75, 1950.

Polanyi, M. (1968). Life's irreducible structure. *Science, 160*, 1308–1312.

Rocha, L. (2001). Evolution with material symbol systems. *Biosystems, 60*(1–3), 95–121.

Schuster, P. (1998). Evolution in an RNA world. In M. G. Ord & L. A. Stocken (Eds.), *Foundations of modern biochemistry, Vol. IV: More landmarks in biochemistry* (pp. 159–198). Stamford: JAI Press.

Schuster, P., Fontana, W., Stadler, P. F., & Hofacker, I. L. (1994). From sequences to shapes and back: A case study in RNA secondary structures. *Proceedings of the Royal Society of London B, 255*, 279–284.

Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Von Neumann, J. (1955). *Mathematical foundations of quantum mechanics* (pp. 418–421). Princeton: Princeton University Press.

Von Neumann, J. (1966). *Theory of self-reproducing Automata* (pp.74–87 and pp. 121–123). Edited and completed by A. W. Burks, Urbana/London: University of Illinois Press. (Original lectures ca. 1952).

Wolynes, P. G., Onuchic, J. N., & Thirumalai, D. (1995). Navigating the folding routes. *Science, 267*, 1.

Zurek, W. H. (1990). *Complexity, entropy, and the physics of information*. Redwood City: Addison-Wesley.

Part III
Afterword - Joanna Rączaszek-Leonardi

Language as a System of Replicable Constraints

For a psychologist interested in language processing, working in the beginning of the 1990s was not at all easy. On one hand, we had at our disposal methods of traditional psycholinguistics, with its information-processing models consisting of symbols, rules, parsers, and mental lexicons. Most of the body of knowledge about language processing gathered since mid-twentieth century was due to research motivated by this approach and its methodology. On the other hand, we were very much aware that the use of language involves time-dependent dynamical processes taking place both within and between individuals and involving physical stimuli, the nature of which, on the first sight, was not obviously symbolic.

The recent (at that time) successes of neural network models of processes, such as word recognition (McClelland and Rumelhart 1981; Seidenberg and McClelland 1989; Kawamoto 1993), emergence of structured semantic representations (Elman 1990) or models of language dysfunctions, such as aphasia (Hinton and Shallice 1991), strengthened the claims that casting explanations in dynamical terms has clear advantages. A hopeful question arose: can we build models of language solely in terms of dynamics, treating expressions of natural language just like any other physical stimuli, instead of endowing them with symbolic properties?

Following this line of thought, it was natural to ask: can *any* cognitive phenomenon be described without referring to symbols but at the same time without endorsing the behaviorist exorcism of mediating mental states? After all, what is a symbol? How is it different from a physical stimulus? Has anybody ever seen one in the brain? Would anything important be lost if explanations of cognitive phenomena were built solely in terms of dynamical, self-organizing brain states that adapt responses to the demands of the environment? When and why would the necessity of a symbol in an explanation of cognitive functioning arise? In what kind of cognitive system? Perhaps it would be easier to ask these questions not about human cognition, which intuitively seems sophisticated and saturated with symbols, but about any living organism that uses memory to adapt to its environment? Do we really need to talk about symbols even at this level? And if yes, what kind of symbols are they?

H.H. Pattee and J. Rączaszek-Leonardi, *LAWS, LANGUAGE and LIFE*, Biosemiotics 7,
DOI 10.1007/978-94-007-5161-3_19,

Led by a naïve faith of a graduate student that some answers are surely just around the corner and can be discovered by talking to experts and reading several books, I started systematic discussions of these issues with Scott Kelso, from the Center for Complex Systems and Brain Sciences (CCS) at Florida Atlantic University, while being inspired and supported in research ideas by other faculty at the Center, most notably Betty Tuller. These scholars made the Center for Complex Systems a home of intense intellectual life and research adventure. At that time CCS was one of the very few places in the world where the complex systems' approach was tested in the domain of psychology, bringing new hope for an alternative framework for studying cognition. At the same time, I was being sobered in my dynamicist zeal by the parallel debates of the same issues with Lewis Shapiro, a Chomskyan psycholinguist, who supervised my dissertation on the processing of ambiguous expressions. Especially during numerous exchanges with Scott Kelso I realized that my questions about psycholinguistics are actually versions of much more basic ones, concerning the bare fundaments of how living things retain and use previous experiences.

Most impressive, from among the literature read and discussed at the time, was the work of physicists that aimed at accounting for informational properties of living matter, such as Robert Rosen, John von Neumann, Alan Turing, and Michael Polanyi. They showed that the quest for the clarification of the nature of intelligence, i.e., the nature of information that enables adaptive functioning of organisms in their environments, had to start at the very beginning of life or at least at the level of very basic principles that make evolution and adaptation possible.

That symbols were thought necessary for these properties was, obviously, not a revelation in the 1990s: after all, the information-processing approach, which emerged 40 years earlier, was based on this claim. But it was not until Kelso suggested that I read the papers of Howard H. Pattee that it became clear that there existed a different interpretation of the necessity of symbols in cognition than the one later embraced by the cognitive sciences. In other words, the symbols that were recognized as necessary in the explanation of adaptive complexity of organisms (e.g., von Neumann's necessity of a self-description (1966)), and in the explanation of human problem solving powers, were mistakenly but all too often identified with the things that computers crunch (see, e.g., Newell and Simon's Physical Symbol System (Newell and Simon 1976)).

Even though alternative, more cautious, ways of conceptualizing information in living systems were already present (e.g., in later papers of Turing (1952), Polanyi (1968), or Rosen (1969, 1991)), Pattee's work was the most comprehensive, in its building from the biological necessity of certain kinds of symbols for control processes to the consequences that doing so would have for memory-based systems in general. Besides, Pattee was able to take a stance informed by important theoretical divides present at the time, and that are still present now. For example, he was active in the discussions between proponents of the information processing approach and of more dynamically oriented "ecological psychology" approach to cognition (Pattee 1982a, b), arguing for the insufficiency of both symbolic and dynamical explanations alone. Perhaps it was this breadth of scope paralleled by the concreteness and precision of the claims that made his work accessible and potentially relevant to the problems we faced in psycholinguistics.

The perspective presented by Pattee changed, and keeps changing, my thinking about language and cognition in general. And I think that Pattee was right in suggesting that it has the potential to forge a third way, alternative to the two dominant approaches in cognitive sciences: the information processing approach and the dynamic approach. This third way is based not on showing that one of the approaches is reducible to the other but on retaining both as complementary models. If this way of thinking about symbols and matter is to be considered, re-constructed, and developed further, it is essential to present biologists, physicists, cognitive scientists and semioticians with the original papers that Howard Pattee wrote from the 1960s to the present. Doing so can provide a solid base from which the hidden assumptions about the role of symbols in living organisms can be made explicit and perhaps questioned within the particular field of study that each of these very different kinds of scientists represent. In the case of the cognitive sciences, this means putting in doubt too uncritical interpretations of the human brain as some kind of computing device, as well as equally too enthusiastic attempts to eschew symbols entirely from our explanations.

In what follows, I will trace the consequences of adopting Pattee's original ideas on the nature of symbolic constraints as informational structures in living organisms for the theory of natural language, and will draw some of its more general consequences for the theories of cognitive processes. Pattee's framework brought several essential shifts in the understanding of very basic terms—such as "symbol," "model," "language," "code," and "communication"—which since the middle of the last century have become crucial in the vocabularies of cognitive scientists and linguists. Accordingly, the first part of this commentary will be devoted to a summary of these shifts from the perspective of a psycholinguist, which means I focus on certain aspects, and elaborate some of them beyond Pattee's work—for example, the importance of the history of physical events for the constitution and workings of symbolic constraints, the coordinative role of the constraints, or the problems encountered while considering referential properties of symbols within a coordinative framework.

Secondly, while Pattee emphasizes that similar basic principles underlie heritable memory structures of a developing cell and language, he also makes it clear that the two informational systems are vastly different. The second part of this paper aims at finding the core similarities and delimiting the range of issues in the theory of natural language that are particularly affected by the abovementioned shifts.

The third section describes the consequences of the application of Pattee's framework to the natural language phenomena. Applying to them concepts enumerated in the first part of the paper means rejecting or loosening some of the assumptions that constitute the basis for many existing theories of language and linguistic functioning. I will show the benefits gained from adopting such a perspective, i.e., the fresh look it affords on the reasons behind ever-recurring problems in linguistic theory, and the suggestions of new directions for theory construction and research that are implied by it. I will also draw attention to the specificity of natural language among other natural informational systems, which will stand out clearly once the common background of basic principles is accepted.

Finally, I will show that the shifts that have to be made are not total discontinuities, and that the traditional ways of looking at natural language can be and should be subsumed within this more encompassing framework—one that actually calls for the parallel and harmonious co-existence of symbolic and dynamic modes of description. In this section I will also venture into the general theory of cognition to expand on the claim expressed above, on the reconciliatory potential of Pattee's work. The most crucial principle of his approach, the necessity of the co-existence of complementary models to describe informationally based processes in nature, indicates a third way, mediating between the two most dominant approaches in cognitive science: the information processing approach and dynamical approach of ecological psychology. **I will argue that this is a way of looking at cognitive phenomena that allows one to be a non-reductionist while remaining a materialist.**

1 Information in Biology: The Complementarity of Dynamics and Symbols

The main arguments of Howard Pattee's approach can be found in his original papers. His commentary upon that work (Part I of this volume) provides his contemporary summary and, at times, qualifications. Thus it might seem that my recapitulation of his main arguments in this section would be redundant. I do not believe this to be the case, however, in that I think it is important to explicate his basic arguments in a slightly different language and from a slightly different angle. This, I think, will foreground these explicit views as well as their unstated implications that have the strongest impact on our thinking about natural language, a topic that Pattee's work is not itself primarily concerned with.

The necessity of using complementary models that is advanced in Pattee's framework leads to seeing in symbols not only their formal properties that have always been appreciated but, above all, their physical nature. In a biological organism, what we call a symbol is always a physical structure that obeys the laws of physics and interacts with the dynamics of a particular system according to these laws. Analyzing this physical nature, the exact physical requirements for something to become a symbol, and the processes that symbolic structures are engaged in, makes on aware that the physical realization is far from being inconsequential for the "formal" properties. This brings about the shift from understanding symbols as "formal entities substituting for something else" to seeing in them "replicable constraints in a particular dynamical system."

Below I will expand on Pattee's postulate of complementarity and the claims about the properties of symbols that follow from this change of view regarding the relation between symbols and dynamics. Among the most important ones are that: (a) symbolic structures are **physical structures**, (b) they are **replicable (transmittable)**; (c) they act as **selected constraints**, i.e., have a history within a system. Further, (d) their constraining role consists in **harnessing dynamics** in a particular way, which (e) is best seen as a functional **coordination** of the parts of the system, be it in morphogenesis or shaping functional behavior. Importantly, (f) such **constraining is continuous**, symbolic constraints may leave a variable role to dynamics: from completely harnessing it to allowing large contextual flexibility.

1.1 Control in Living Systems Requires Complementary Models

Pattee's work was initially concerned with information in living systems at a very basic level—i.e., analyzing how systems that obey physical laws can undergo evolution. This investigation went in two directions: one was to spell out the physical conditions for the emergence of memory structures capable of ratcheting the adaptive complexity of an organism in the face of environmental challenges (e.g., Pattee 1968). The second was an attempt to answer an old question about the sufficiency of the physical laws to account for such memory structures (Pattee 1969). The outcome of this work was Pattee's conclusion that two complementary models are necessary for the explanation of informational processes in living organisms, along with a list of necessary physical conditions needed for memory structures to emerge.

Pattee treats complementarity as a very basic and universal epistemic necessity, always present when a functional reduction of degrees of freedom is made and recorded in a system. He addresses specific examples, including the contexts in which this notion usually appears in the scientific literature, such as the distinction between laws and initial conditions, and the complementarity principle in quantum mechanics. Other authors point to self-organization and emergent phenomena as requiring more than one mode (or rather level) of description (e.g., Kugler and Turvey 1988). However, the complementarity that is relevant for the study of natural languages can be recognized without descending to the level of quantum description—and for reasons different than the appearance of a novel level in a self-organizing system.

The impossibility of using just one model rests in the fact that the constraining role of a memory structure within an organism depends on something more than just the laws of physics alone: it depends also on its history within a particular system, which includes the irreversible and probabilistic process of natural selection. The way in which a memory structure constrains dynamics is selected on the basis of the adaptive properties of the phenotype (i.e., the effects of the constraints' action) and the transmittability of that structure. At any point in time, the dynamics itself, as a reversible process, can be described by the laws of physics. *How* it is and has come to be constrained by the selected structures to bring about adaptive functionality in an environment will, however, escape this description.

Pattee's work elaborates on the views of two prominent figures of the time: Michael Polanyi and John von Neumann. Polanyi, in his *Life's Irreducible Structure* (1968) points out that any system that harnesses natural dynamics to perform *useful* work, be it a machine or a living system, is under "dual control." On one hand, it has to conform to the laws of physics, but at the same time it possesses a structure, a design, which provides the boundary conditions within which these dynamics work.

> (…) if the structure of living things is a set of boundary conditions, this structure is extraneous to the laws of physics and chemistry which the organism is harnessing. Thus the morphology of living things transcends the laws of physics and chemistry. Polanyi (1968, p. 1309)

Laws of inanimate matter can thus serve as explanations of systems *within* their existing boundary conditions (including self-organization phenomena). However, the boundary conditions themselves require a complementary description. This claim is perceived by Polanyi as liberating, rather than limiting: "*recognition of the impossibility of understanding living things in terms of physics and chemistry, far from setting limits to our understanding of life, will guide it to the right direction*" (p. 1312).

Polanyi talks about "alternative" description—but does not specify *how* this description differs from ones formed in terms of physical and chemical laws. The proposal of "symbols" as terms of this description appears in Pattee's papers together with appreciation of the work by John von Neumann and other theorists of computation. Von Neumann advocated the logical necessity of the existence of a separate structure in the case of evolvable organisms, i.e., "that information in the form of non-dynamic symbolic constraints ("quiescent" descriptions) must be distinguished from the construction dynamics they control in order to allow open-ended evolution." (von Neumann, after Pattee (2006, p. 225)). It is also important to note, for the later discussion of generalized and natural language, that in his theory of self-replicated automata, von Neumann has demonstrated that there is a threshold of complexity below which sustaining such transmittable structures is not possible (von Neumann 1966).

1.2 Rethinking the Nature of Symbols

The existence of such potentially functional (or meaningful) structures in an organism is thus seen as a necessary condition for the adaptive increase in complexity. On the other hand, it is admitted that accounting for their functionality requires a different description than can be given in terms of laws of physics. **A physical structure becomes a boundary condition because of its history, including natural selection**. These non-dynamic (with respect to the current dynamics of a system) and replicable structures Pattee calls "symbolic" constraints, and their systemic structure—a "language" (e.g., Pattee 1972). They are discrete, reproducible (transmittable) and they serve as the instructions to re-construct a system (von Neumann 1966). However, their inseparability from and reliance on dynamics makes them quite different from how formal symbols are usually understood. In what follows, I elaborate on the essential features of such "replicable physical constraints," concentrating on those that contrast with "symbols" as the term is usually understood—i.e., as elements of purely formal systems.

1.2.1 Symbolic Structures Are Physical

The informational structures in biological organisms, even though their role requires an alternative description to the ones posited in terms of the laws of physics, nevertheless are not abstract: they remain physical structures.

> (…) all forms of codes, rules, or descriptions, even the most abstract and symbolic, must have a definite structural basis. Pattee (1973a, p. 101)

Thinking about symbols not as abstract entities but, rather, as physical constraints has many consequences; let me mention two that seem the most profound: (i) being physical, they can be in a non-mysterious way causally related to the processes they control; (ii) both their workings and their copying have to obey physical laws and have to be stable and reliable under these laws. Both claims put conditions on the apparent "arbitrariness" of symbols as carriers of replicable constraints.

It is very important to note that the necessity of dual description does not imply a Cartesian dualism: all the constraints on dynamics ARE physical structures, obeying the laws of physics. What makes a molecule a message (i.e., endows it with the power of physically harnessing the dynamics) is not its particular physical or chemical properties, but its evolutionary history within a system: i.e., its being selected for bringing about a particular phenotypic effect. What makes an interaction between molecules "executing a constraint," i.e., something more than physical interaction or collision, is the selection of the effects of such an interaction, based on the previous history of a structure within a system. *This* is what makes the constraints "meaningful."

> We are taught more and more to accept the genetic instructions as nothing but ordinary macromolecules, and to forget the integrated constraints that endow what otherwise would indeed be ordinary molecules with their symbolic properties. Pattee (1972, p. 249)

1.2.2 Symbolic Structures Are Replicable (Transmittable)

In order for a physical structure to be selected as a constraint and to further serve as a memory, it has to be stable enough to be copied, and its replication should be relatively energetically cheap. Transmittability means that symbolic structures in living organisms undergo two different, independent processes of "reading": (1) as controls, i.e., functional constraints on dynamics and (2) as structures to be copied. In the second process, the functional "meaning" of a constraint should not matter, while the physical structure itself undergoes replication.[1] It is important to note that both processes are sources of selection criteria determining the shape of symbolic structures (i.e., those are selected that constrain functionally and that transmit well).

---

[1] However, the value of a constraint in a system may be linked to the reliability of copying (i.e., copying of particularly important constraints may be additionally warranted by error-correcting mechanisms).

1.2.3 Constraints Are Selected

In order for these structures to be "informative," i.e., being able to send the dynamics on an adaptively functional path, at the moment of their action they must be physically indeterminable. If we take as an example the informative properties of DNA, the crucial thing is that it is highly improbable that its physical structure, e.g., the order of the bases, would appear by natural laws in a single given moment. **Its presence in a system and its shape may be explained only by reference to a process that takes place on a different timescale than that of its current, constraining action.** Specifically, it is the process of the natural selection of successful phenotypic forms, which arose under the control of such informational structures, that explains the functional nature of a current structure. In Polanyi's words, it is the very improbability of its occurrence that enables such a structure to be an information structure (Polanyi 1968).

The improbability of a structure in a given moment determines its informational value (according to Shannon's definition of information), but, obviously, does not say anything about the meaning of such a structure. As Polanyi says: "*(…) the improbability count gives the* possible*, rather than the* actual *information (…)*" (p. 1309). The actual meaning of the structure is the construction, under its control, of the morphology of the offspring, functionally adapted to its environment. Thus "how informative a structure" is (the quantity of information) does not have a straightforward relation to its meaningfulness, i.e., the epigenetic construction of shape or behavior.

Selection is a **historical** process that underlies current function. The particular history of a symbolic structure in a particular system is crucial for determining its meaning. A constraint will behave just as would any other physical structure on which physical laws act here and now. What is meaningful is its being in a particular place at a particular time, and that is historically determined. One of the crucial conditions for selection is variability—thus the capacity of low-cost variation in structures is an important feature of potential informational structures.

1.2.4 The Constraining Role of an Informational Structure Consists in Harnessing Dynamics

Symbols do not "carry" any meaning by themselves, nor do they, in any easy way, map to or "represent" external world: "*It is useless to search for meaning in symbols without complementary knowledge of the dynamics being constrained by the symbols*" Pattee (1987 p. 337). Symbols harness a system's existing dynamics so as to generate a specific structure or behavior. The meaning of such a symbolic structure is what it *does* with these dynamics. Thus, in its evolution and selection, such a structure "relies" on natural dynamics to construct functional structures or behaviors. This is why so little information can control such complex construction processes. Even more importantly, this is why it is misleading to talk about transmittable constraints as "representations" or "models" of the external world.

Symbolic structures are, rather, **memories of the choices** of adaptive coordinations with reality, of constraints on existing dynamics that have led to adaptive results. They are "memories" of past decisions made at systems' dynamic bifurcations. Understanding symbolic structures as constraints on existing dynamics unburdens them from the role of being the sole meaning-carriers. The important part in realizing meaning falls on dynamics. This, in turn, assures that informational constraints are flexible, and that the outcome of their action is adapted to a concrete system and situation. Forces active in a specific physical system and situation in a given moment are part of the meaning construction process.

Another important implication of this understanding is that since symbols arise from and control dynamics, the causes for symbols' structures should be sought in these dynamics and not (or not only) in a system of rules existing somewhere independently of them. Analyzing such systems in a structuralist manner may uncover interesting regularities, but such description of these systems is just that: a description. Without accounting for the underlying dynamics, it cannot constitute an explanation (Rączaszek-Leonardi 2009b). Thus, in the study of a system of symbols, it becomes of primary importance to uncover the types, timescales and systems within which the relevant dynamics take place.

1.2.5 Constraints Reduce (Coordinate) Degrees of Freedom

Constraining a dynamical system of many parts in a functional way requires the selection of its relevant degrees of freedom so that it can generate an adaptive structure or perform an adequate behavior. This, in behavioral sciences, is termed "coordination" (Turvey 1990). Seeing the symbolic role as coordinative brings a shift in our understanding of communicative processes in the direction of coordination, rather than transfer of information. In physics, the term "coordination" is typically used to describe temporal coherence. In this context, however, selection processes assure functionality—thus, coordination through symbolic constraints imposes **functional temporal coherence** on a system that has its own natural dynamics.

1.2.6 Constraining Is Gradual

Functional constraints may play a variable role in constructing a structure or organizing behavior, relative to natural dynamics. Constraints may harness the dynamics completely (which probably rarely happens in biological systems) or leave more space for its role, thus making the outcome less predictable and more contextually flexible. Constraining is thus not a 0-1 process, but may be characterized by degrees of exerting a specific influence on dynamics.

The properties of symbols listed above are necessary for any physical system to retain and to pass on information. I chose to describe those characteristics here, because they are incongruent with the traditional views on symbols, and because they will be particularly important for thinking about natural language. However, before I turn to that discussion, the grounds for making such generalizations have to be specified.

2 Delimiting the Scope of Generalization

Pattee's initial work, as mentioned earlier, was geared towards seeking the principles underlying the origins and transmission of informational control in biological systems. This work had consequences for the philosophy of biology, positioning this science in relation to the physical sciences, in its tackling of problems such as physical preconditions of information generation and transmission in living organisms, on the one hand, and the limits of description in terms of physical laws on the other. Pattee himself initially seemed most interested in the consequences of his work for the origin of life problem. Later however, since the 1980s, he developed his ideas on the applicability of the principle of complementary modes to cognitive and cultural systems. It is important to see why this framework can be applied to such different and evolutionarily distant systems of symbols.

2.1 On symbols Tame and Wild: Natural Language is not a Formal System of Symbols

Above, I listed those characteristics of the symbolic system of constraints in a living organism that require us to rethink our stereotypical notions about symbols. It is clear that the properties of symbols and their meanings derive from the reliance on dynamics that the constraints evolve with. As such, these symbol systems do not seem to be easily translatable into formal symbol systems, that can be transformed by syntactic rules without reference to their meaning. What then could their relation to formal systems be? One might see formal symbol systems as a subtype of this more general class—a subtype, in which the dynamics are artificially reduced to the point of being inconsequential.[2] In other words, formal symbols do not rely on lawful dynamics for their meaning; in fact, they do not *have* a meaning, until one is externally given.

Now, a question arises: How justified was the mid-twentieth century claim that both cognitive and linguistic processes were describable in terms of (or reducible to) this narrow subclass of symbolic systems? The computer metaphor for human cognition meant just this: judging cognitive processes to be computable on any kind of machine (e.g., Turing's universal machine) meant casting those processes in terms of computable formal symbols. Yet wasn't this an approach dictated by the fact that these were the symbols most familiar to our thinking? Or was the approach dictated by practical reasons—i.e., by the availability of devices that can process such symbols? In Pattee's words:

> (…) I have argued that the most fundamental concept of a constraint in physics depends on an alternative description, and that the apparent simplicity of constraints is in fact a property of the language in which it is described. Pattee (1972, p. 248)

---

[2] By this I mean inconsequential within the system. The dynamics that makes transformations within such a system meaningful is, artificially, pushed out into the larger system—i.e., that of a person using the formal system.

In other words, it is agreed that explanations of cognitive and linguistic phenomena need symbols. These processes are claimed to have symbolic properties that require alternative descriptions to the ones in terms of laws of physics, but describing them as formal symbolic systems obscures their true nature.

Formal systems require abstracting from dynamics, an "overdetermination" of the result, because they arose for very specific human purposes. This kind of formal precision is not required for cognition in the wild. On the contrary, context-insensitivity (stemming from "cutting off" the dynamics) and "exactness" of the processing, determined only by the form, would most probably prove deadly—leading to the inflexibility of a living system, and its helplessness in the face of error. Moreover, from the theoretical angle, not acknowledging dynamics puts the burden of explanation exclusively on formal symbols. Reconstructing the dynamics in explanatory models in terms of formal symbols becomes cumbersome and unnecessary.[3]

Making the conceptual shifts in understanding symbols in our thinking about natural language brings both theoretical and methodological advantages. Language, unlike other aspects of cognition has an intuitively clear symbolic level, and thus is especially prone to the attempts of "fast formalization." In fact, one of the key factors that brought about the "cognitive revolution" was the conceptualization of language as generated by a formal system, and therefore as a phenomenon completely describable on the symbolic level (Chomsky 1957).

On the other hand, by the very same virtue of being based on distinguishable symbol-like, sequential entities, language shows affinities with the genetic information structure in living organisms (e.g., Jacob et al. 1968, Bernstein 1965, as quoted in Jakobson 1989, pp. 442, 444). In fact, the metaphor of "language" in reference to the workings of genes appeared both in biology and in linguistics (Jakobson 1989, pp. 442–446), most intensively after the discovery of the mechanisms behind biological information transmission by DNA. Sometimes, however, it is not clear what *kind* of "language" is supposed to have this "strict analogy" to the genetic information system: is it the natural human language, the formalized version of that language, or an abstract formal language? If it is natural language, what properties justify this analogy?

For Pattee, as he states in his commentary, the use of the term "language" to address biological information processes was quite natural and consistent with the *Zeitgeist* of the time. As I understand the scope and intension of the term, on the basis of his papers written over several decades, he delimits a certain very basic phenomenon that is assumed to be **the same** both in the system of information transmission in biology, and in natural language. This core "rudiments of a theory of symbol systems" (Pattee 1980), is the property that enables the transmission of useful constraints on dynamics. Thus, the term "language" pertains to all phenomena that are based on this general relation of physical system of transmittable structures to the dynamics around which it arose:

> My approach is to generalize measurement and linguistic functions by examining both the most highly evolved cognitive systems and the simplest living systems that are known to have the potential to evolve, and abstracting their essential and common measurement and

---

[3] A good example is the attempt to express dynamics present in the external world as a set of discrete stimuli in the early theories of perception (e.g. feature detection theory). Gibson's theory of perception was a reaction to such attempts and a way to let dynamics back in (e.g., Gibson 1960, 1966).

> linguistic properties. I want to emphasize that when I speak of molecular language strings and molecular measuring devices I am not constructing a metaphor. Quite the contrary, I mean to show that our most highly evolved languages and measuring devices are only very specialized and largely arbitrary realizations of much simpler and more universal functional principles by which we should define languages and measurements. Pattee (1985, p. 268).

By giving this minimal, but specific, intension to the term, Pattee, like the abovementioned biologists and linguists, allows its extension to include phenomena that few contemporary linguists would call a language. The term is thus applied to biological phenomena in a literal, non-metaphoric sense: the claim is *not* that biological information transmission is (in some respects) like the human language—but, rather, that both cultural and biological information transmission require a generalized language, and that some basic element that exists in each of them is exactly the same.

This insight has at least two important terminological consequences: (1) that in order to talk about specific instances of such constraint-transmission systems, an additional adjective has to be used, as in: biological (genetic) language, natural (human) language, formal language, and (2) that the term "language," in all its meanings, encompasses both the symbolic constraints *and* the dynamics that are constrained—the two aspects are inseparable if one is to understand the function of language in a system. This latter consequence heralds an important change from the traditional use of the term "language" in reference to human language—which, as mentioned earlier, most often indicated only the symbolic layer of this human communication system.[4]

Accepting Pattee's general principle of complementarity, too, means agreeing that it does not make sense to talk about symbols without talking about the dynamics that maintain symbolic structures and, on the other hand, are harnessed or controlled by them. In this context, using the term "language" emphasizes the necessity of rate-independent structures in biological systems—and this, contrary to the contemporary general propensity to describe biological structures in physico-chemical terms, brings attention to the *symbolic side of physical systems*. When addressing the theories of natural language, however, accepting the same principle leads to the emphasis of the *dynamic side of the symbolic system*. In other words, Pattee's view is that particular properties of replicable physical constraints in living organisms make those constraints symbolic, and their system a language. Here, on the contrary—or rather, complementarily— I will underscore that certain properties of symbols in the natural language system make them transmittable physical constraints on dynamics.

Now, obviously, any comparisons between such evolutionarily distant systems as biological heredity and natural languages should be made with caution. Human language is surely a very different phenomenon than other types of informational processes in living matter (see also Pattee's commentary for the discussion). Its symbols are certainly elements of a different system, with different properties

---

[4] The term "languaging" recently has appeared in the language sciences, adopted from the works of Maturana (1978), see also Cowley (2012). Its increasingly frequent use may testify to the pressures for including dynamics in our explanations of the human language system.

(e.g., different copying mechanisms) and different forces shaping their structure. To repeat: acknowledging the immensity of differences, I assume, after Pattee, that **an essential feature common to these systems is the specific relation between symbolic and dynamic modes**. In other words, both systems are examples of "symbols in the wild" forming linguistic systems crucial for natural information processes. And in the case of both one should be careful when assuming equivalence with artificially created, "tame," dynamics-free, formal symbols.

Below I will show that emphasizing the necessity of co-existence of symbols and dynamics in natural language has important theoretical and methodological consequences: On one hand, it gives a new perspective on recurrent problems in linguistic theory, such as: Why is it so notoriously difficult to specify a complete grammar for a language? How can one account for the context-sensitivity of linguistic expressions? How is it possible that natural language is so effective and so flexible? I hope to be able to show that the problems that traditional linguistic theories have with these questions are due to artificially cutting off the symbolic layer of linguistic structures from the natural dynamics that they arose to constrain, and their history in the dynamical system during which they were selected.

On the other hand, however, giving this "explanatory share" to dynamics obliges the researcher to specify the kinds and timescales of the relevant processes. This is certainly not an easy task. Studying dynamics requires employing new methods that are designed to deal with complex dynamical systems, and the analysis of the time course of events, stability, degrees of synergy, type and strength of coupling, etc. This task, I believe is already under way, visible especially in the work on language perception and production (Fowler 1980; Kelso et al. 1984; Tuller et al. 1994), language evolution supported by computer simulations (Steels and Belpaeme 2005; Smith et al. 2003) and in some work in psycholinguistics (MacWhinney 2005; Rączaszek-Leonardi and Kelso 2008; Rączaszek-Leonardi 2010, and others mentioned in Section 3 below). However it is also true that, compared with the influence on research paradigms exerted by the traditional psycholinguistics, this approach is still in its infancy.

2.2 *On coding and Meaning: Natural Language is not a Code*

Given that there is a variety of metaphors and comparisons for the workings of genes that use as a source the workings of mental processes (such as memory, code, language), it is important to very carefully and precisely point to the property, proposed here to be shared by genes and languages, that is at the core of informational processes in living systems. Is this core "linguistic" relation, then, the relation of "coding"? Shall one talk about the "genetic language" or "genetic code"? Or both? And if both, how are they different?

The most intuitive way of drawing an analogy between DNA and language has its basis in understanding natural language as a code—i.e., as a system of forms "standing for" their meanings. Accordingly, certain combinations of nucleotide bases of DNA were seen as standing for certain functions and processes in a cell (or even, before the advent of the epigenetic approach, certain features of a phenotype), just as certain combinations of linguistic signs were seen as standing for certain concepts, sets of semantic features (e.g., Katz and Fodor 1963), or sets of referents. Such a view of language is possible because of the reification of meaning as something stable, amenable to a static and discrete description (e.g., an intension or extension of a concept corresponding to a word) that can be mapped to another static structure (a sign).[5] It is also congruent with the so-called "conduit metaphor of communication" (Reddy 1979), i.e., communicating as sending something (signs carrying reified meanings).

While the expression "genetic code" does not seem controversial, proposing "language" in the workings of genes has raised doubts (e.g., Oyama 2000). Yet it seems crucial to note that the common property that Pattee claims to be at the core of informational processes is the specific **relation of symbols to dynamics**. Rather than the relation of *mapping*, that pertains to "codes," this relation is of constraining, that pertains to generalized "languages." It is the constraining relation that is the *meaning* relation. In other words, "linguistic" in Pattee's sense does not mean "code-like," whether in reference to biological or natural language processes. This important distinction may be blurred, because the two relations were understood as being equivalent by many theories: "The view of linguistic communication as achieved by encoding thoughts in sounds is so entrenched in Western culture that it has become hard to see it as a hypothesis rather than a fact" (Sperber and Wilson 1986, p. 6).

Consider the following example, which posits that a message (chosen by a speaker for linguistically irrelevant reasons) is "*encoded in the form of a phonetic representation of an utterance by means of the system of linguistic rules with which the speaker is equipped. This encoding then becomes a signal to the speaker's articulatory organs, and he vocalizes an utterance of the proper phonetic shape. This is, in turn, picked up by the hearer's auditory organs. The speech sounds that stimulate these organs are then converted into a neural signal from which a phonetic representation equivalent to the one into which the speaker encoded his message is obtained. This representation is decoded into a representation of the same message that the speaker originally chose to convey by the hearer's equivalent system of linguistic rules. Hence, because the hearer employs the same system of rules to decode that the speaker employs to encode, an instance of successful linguistic communication occurs.*" (Katz 1966, p. 103–104).

However, once one takes the Pattee's view (that symbolic structures are constraints on dynamics), it becomes clear that symbolic constraints **do *not* "map like a code" to the consequences of their actions**.[6] Effects of constraining are naturally context-dependent (crucially relying on the dynamics being constrained), thus are

---

[5] The semiotic shift which is based on recognizing that the relations between a sign and its referent is a 3-element relation, i.e., one that includes the interpretant that contextualizes the reference, seems to do part of the job in "uncodifying" the meaning relations.

[6] Pattee calls these constraints 'referents': "The referent of a symbol is an action or constraint that actually functions in the dynamical, real-time sense. Here is where any formal language theory loses contact with real languages." (Pattee 1980, p. 263). However, I will refrain from using the notion of referent, in order to avoid the reification of a symbol's action.

predictable only to some degree—and, in organisms, are always underdetermined. In natural language too, it is this constraining relation that is the *meaning* relation; and it can arise only from the history of a certain physical structure as a constraint on certain system's dynamics in a certain environment. Such a key property is simply absent from the code-view of language; thus in order to acknowledge the crucial equivalence of the natural information systems, this view of language has to be abandoned. This has been already recognized by numerous theorists of language, among them those who undermine the apparent simplicity of the relation between form and meaning, grounding it in the dynamics of interaction (Merleau-Ponty 1960; Cowley and Love 2006; Kravchenko 2007).

Now, does this mean eschewing the notion of coding relations entirely from the explanations of natural language or the explanations of information in biological systems? Obviously not. It only means that the relation of "coding" and the relation of "meaning" have to be clearly distinguished.[7] To code is to map one symbolic structure onto another symbolic structure. Meaning is a relation in which a symbolic structure acts to harness dynamics, endowed with this power by the process of natural selection within a given system.

The clear correspondence of the sequence of nucleotide bases to the sequence of amino acids seems to be a good example of a code. So does the correspondence between dots and dashes in a Morse code and letters of alphabet. Both are different from meaning relations, which are based on constraining dynamics in a functional way. In other words, symbols can be coded in another set of symbols, perhaps for a better adaptation to a given transmission medium (e.g., the Morse code is better adapted to a telegraph than the alphabet) but it does not make them more, or less meaningful. A code is not a language.[8] It thus remains an open question as to whether or not the coding relation is at all necessary for a system of symbols to be meaningful. It does not seem so, as long as the system of symbols is itself replicable and capable of variability, and thus can be subjected to natural selection.

Spoken language seems to be just such a system of meaningful replicable constraints: spoken expressions have meaning, as they constrain interpersonal

---

[7] It remains to be seen if both can be subsumed by some more general concept—i.e., of a relation that may hold both for mapping between forms *and* being a potential constraint. Perhaps the constraining relations could be treated as a more general concept. Coding would then be seen as a special case of 'superconstraining', to the point of becoming mapping. On the other hand, a very general definition for 'coding'—such as, e.g., used by Barbieri, that coding is setting "*the rules of correspondence between two independent worlds*" (Barbieri 2003, p. 94), might, somehow, encompass the flexible relations of constraining, making it just another type of code. The latter possibility seems implausible though, given the dynamic and historical nature of the constraining effects. For now, let me set this discussion aside and, for convenience, distinguish between the two types of relations.

[8] Similarly the Braille code: Although numerous petitions arrive at the Braille Authority of North America (mainly from students who would like to count a Braille course towards their second language requirement), the Position Statement of BANA issued in 2008 is: "Braille is not a language but a code." And further: "To call Braille a language would be comparable to calling 'print' a language."

and individual (on-line and ontogenetic) dynamics in a reliable way. They do not have to be written down in an alphabet in order to mean; nor they are a code themselves—i.e., they do not map onto some kind of internal symbolic structure, a "mentalese." Coding processes in natural language include transforming phonological realizations of expressions to written forms, letters of alphabet or ideograms. Such coding is a remedy for the transiency of human speech, an adaptation to a different mode of transmission. Codes can be further coded, as when an alphabet is transformed into a Morse code or Braille. This further changes their potential for transmission, adapting to different media or modalities, but does not change their meaning. Adaptation to transmission media means making the symbolic system more stable under the process of copying (i.e., copying is more accurate), or making the copying process less energy consuming.

Similarly, both the relation of coding and of meaning can be identified in biological informational systems: the "genetic code" is the—stable, fixed—relation of nucleotide bases to amino acids, where the meaning of the symbolic structures of DNA is in the constraining role the proteins have with respect to the cell's dynamics. As Pattee sees it, while the DNA bases code for the amino acids, it is the *folded* amino acid sequence (the protein enzyme) where the first informational constraint on dynamics occurs. "*Folding transforms what are essentially rate-independent syntactically coded sequences into rate-dependent functional controls*." (Pattee 2007, p. 10).

Thus both the *meaning* and the *coding* relations are present in informational systems. Yet it seems that great confusion between them takes place, especially in our explanations about natural language. Linguistic expressions are often seen as forms that "map to meanings"; similarly, some initial theories of heredity proposed that genes map on, code for, or even "contain" phenotypic traits. In order to effectively "map" forms to meanings, the events in both domains should be clearly defined and individuated. Doing so is often difficult in the cases of genetic actions and linguistic actions. In natural language, writing is a code for spoken expressions, but it is the spoken expressions that are *the* level at which meaning relation should be sought. In a cell, DNA bases are the code for amino acids, but it is the protein enzymes which interact physically with the dynamics of the cell and which is *the* level at which the function or meaning relation should be sought. Paradoxically, language conceived as a code thus lacks the generalized "linguistic" property, which is based on symbols constraining dynamics. But below I will show how endowing natural language with the above understanding of symbols, i.e., making it a linguistic system in a generalized sense, provides a framework which opens more productive ways for studying it.

3 Language as a System of Selected, Replicable Constraints

This section shows natural languages as having the essential properties that make them, according to Pattee, carriers of functional constraints. Other work has put emphasis on how to introduce dynamics in the study of language both theoretically

(Rączaszek-Leonardi and Kelso 2008; Thibault 2004) and empirically (Kelso et al. 1984; Fowler and Saltzman 1993; Tuller et al. 1994; Cowley 2004; Shockley et al. 2003; Fusaroli and Tylén 2012). The importance of the dynamics as a source of language structuring has been also investigated, see e.g., Smith, et al. 2003; Rączaszek-Leonardi 2009a; Lupyan and Dale 2010, with the concurrent analysis of the role of psycholinguistic data for the theory of language (Rączaszek-Leonardi 2009b) and methodology for psycholinguistic research (Rączaszek-Leonardi 2010). Here, having these developments in mind, I will spell out in a systematic way the reasons why natural language symbols should not be conceptualized as formal symbols. First, the shifts in thinking about symbols described above are applied to natural language symbols. Then the subsequent sections describe some of the implications of such a turn, which undermines the view of language as an individualistic referential system of arbitrary symbols and points to fruitful ways to study it.

3.1 Soundwaves with a History

Below, each property listed in I.2 is related to natural language symbols to show the advantages of treating them as replicable constraints.

3.1.1 Natural Language Symbols Are Physical

> We operate with our own natural language as if it needed no structural basis whatever, and when we speak of other types of symbol systems we usually carryover this unjustified abstraction. Pattee (1973b, p. 143).

Pattee rightly points out that the physicality of natural language stimuli is especially prone to be neglected, and therefore seen as inconsequential. The task of designing a theory of the physical realization of linguistic expressions is usually handed down to phonetics, and separated from syntax and semantics—the sciences of structure and meaning. However, without physical instantiation, symbols cannot have any causal powers. The causal power of a word can be acquired only by the participation of its physical realizations in innumerable interactions. In other words, natural language symbols are capable of evoking certain meanings because they participate, as physical stimuli, in various forms of social life. In such social situations, they are strong physical stimuli (most often verbal actions, embedded in other actions) capable of influencing the coordinative situation and modifying it. Among other criteria, it is for the effectiveness and functionality of this modification (for higher-level coordinative aims) that the symbols are selected.

Accepting the view of symbols as constraints makes this physicality stand out. It makes one aware that a physical symbol always co-occurs with, is only a part of, other physical events. Physically, most often, it is a soundwave produced in an interaction, but it is crucial to note that it is a vocal tract gesture that results in a soundwave and that can be apprehended through it (Browman and Goldstein 1989;

Goldstein and Fowler 2003). The mystery of the causal power of words thus appears less deep: speakers do not act with abstract, immaterial signs. They act, producing "soundwaves with a history." Now, in order for those soundwaves to have acquired functionality, i.e., to have a meaningful history, they have to be stable and reliable under the laws of physics and principles of perception, and they have to have the power to evoke the desired effect. As I said earlier, the requirements of particular media may make it necessary to code the physical symbolic structures into other physical forms (e.g., writing or a binary code). This however does not make the "original" symbols any less physical.

Thus, what makes a natural language symbol a message, i.e., what makes it different from any other soundwave produced by a human, is that it has been selected and stabilized in the process of cultural evolution; paraphrasing Pattee (1972, p. 249) quoted above, (p. 301), these are the "integrated constraints that endow what otherwise would indeed be ordinary [sounds] with their symbolic properties."

3.1.2 Natural Language Symbols Are Replicable

In order for a physical structure to be selected as a constraint, and to further serve as a memory, it has to be stable enough to be transmitted and its replication should be relatively energetically cheap. This puts conditions on the arbitrariness of linguistic expressions: they are formed according to principles of least production effort and greatest perceptual distinctness (Lindblom et al. 1984; Oudeyer 2006). These conditions influence both the forms (the phonetic structure of utterances) and their larger structures (e.g., sentence length may be correlated with mean utterance length, constrained by memory requirements; compositional structure may arise under the pressures of learnability (Smith et al. 2003)). The system is thus adapted both to the medium, as well as to the capabilities of the users operating in this medium (Deacon 1997). Recent research in the emerging field of experimental semiotics shows how the forms and structures of symbol systems artificially designed for collaboration are structured by various demands and properties of the medium (Fay et al. 2010; Galantucci et al. 2010).

3.1.3 Symbols Are Selected Constraints

As noted in (a), above, what makes physical stimuli the carriers of constraints is their history in the system. In natural language, this history involves processes on several timescales: a structure is selected for bringing about a particular effect (here and now; for interaction or an individual) and for being transmittable (both here and now: heard, perceived; and in ontogeny: learned). In the cultural timescale processes that maintain the coherence of the linguistic systems, the compatibility and relevance of the constraints brought by a structure with respect to other structures are decisive. The form and structure of expressions of natural language are thus shaped by these multiple selective forces. In this context, studies of the structural properties of symbols in natural language should include both the diachronic analyses of the use of language and the data on language development and learning (Tabor 1994; Lupyan and Dale 2010; Smith et al. 2003). Also, it is important to remember that for selection to occur, variability is needed. The awareness of these facts makes a researcher look differently at the so-called “nonstandard” uses of language. Rather than imperfections of processing, they are seen as variations necessary for the plasticity of a system (Steels 2006), which “seeks” to find a better fit to the coordination goals in the process of cultural evolution. Non-standard uses can thus be seen not only as errors (which obviously occur), but also as constantly “trying on” new structures for various coordinative functions.

Obviously, the fact that linguistic symbols are physical structures immersed in many other processes is not something that a speaker needs to be aware of. On the contrary: it may well be that the individual phenomenal experience of a symbol’s stability, and even, perhaps, that of a clear reference, is needed for it to perform a coordinative function at the level of interaction.

3.1.4 Linguistic Expressions Are Constraints that Harness Dynamics.

Meaning arises only through the behavior of constrained dynamics. Linguistic symbols do not “carry,” or “transfer” meaning. Seen from this perspective, symbols of natural language do not map to or “represent” the external world; the meaning, or what might be just an element constituting “meaning” in such a complex view, rests in how they are able to change the existing dynamics of a system composed of interacting individuals. Again, this is far from a code-view of language. Thus, both harnessing individual cognitive dynamics and constraining the dynamics of ongoing interaction constitute the meaning of linguistic expressions. Even if effects for the cognition of an individual are often seen as being of primary importance (e.g., Chomsky 2011), the harnessing power is gained in the processes of selection and one of the most important selection criteria for a structure is its **effectiveness in inter-individual coordination**. The conviction about the primacy of this social-coordinative function of language is obviously not a novelty: “*(…) each higher function initially had been a particular form of social collaboration and only later it transformed into individual behaviour, interiorizing structures (…).*” (Wygotsky 1930/2006, p. 62), the transformation being due to internalizing language.

Therefore it would be difficult to maintain that meaning can be found in conceptual/semantic networks of individual minds or in objects in the external world. Meaning is what an expression does in a situation. Such an approach turns language investigations in the direction of functionally and pragmatically oriented theories of language. The structuralist and poststructuralist methodology of looking at the superficial systematicities of the selected forms will tell us close to nothing about how the structures arose and therefore what they might do in various situations. As can be inferred from the above discussion, proposing a generative machinery in the language user’s mind also will not do the job. Studying only the symbolic mode of natural language makes little sense without a complementary study of dynamics in which the symbols are immersed.

Thus, the study of language should be the study of forms in their function of constraining dynamics. Reliance on dynamics gives the utterances context-sensitivity "for free": the utterances are (almost) always contextually relevant, simply because context (the existing dynamics) is always an essential part of the meaning (Rączaszek-Leonardi and Kelso 2008). Symbols contribute only how dynamics is to be constrained. This also explains the effectiveness of language, i.e., the fact that the same expression may mean infinitely many things, depending on the context (e.g., Barwise and Perry 1983), i.e., on the dynamics in which it is immersed.

If we agree that symbols do not carry meanings by themselves, then instead of thinking that the symbol's meaning is disambiguated by situation and previous expressions, we can say that it is rather an utterance that directs the fate of an interaction and thus chooses one of its many possible developments, a path which will be taken by a system of conversing people. In this approach, the common ground, understood as shared dynamics (physical as well as cognitive) is already present in each interaction and further shaped and stabilized by language. Instead of being a background against which symbols are exchanged, such common ground is an essential part, a substrate of meaning to which symbols bring only small "pushes", making it functional in a situation. Looking at linguistic interaction without the "written language bias" (Linell 2005) reveals that language does not constitute interactions but rather constrains existing ones—in accordance with the view that a "*good biological as well as good engineering design makes the maximum use of natural (non-informational) constraints and laws of nature, so that the control information can be kept to a minimum*" (Pattee 1982a). Such principles of reliance on the inherent dynamics of a system, with external constraints providing "only" a functional binding of the degrees of freedom of a system, have been already used in behavioral sciences, e.g., at the level of motor control (Bernstein 1967).

Including the dynamics as an essential element of explanation takes the explanatory burden off the linguistic structures, but, on the other hand, obliges researchers to identify the timescales and systems in which the relevant dynamical events unfold. This means going back from studying mainly written language or grammaticality judgments of single sentences to the real-life linguistic interactions (Schegloff et al. 1996). With the possibilities of the data gathering and analysis we have now, it becomes feasible to see utterances as a part of on-going interactions, analyzing these on the pico-scale level (Steffensen and Cowley 2010) but also tracing the interaction dynamics, both on the level of physical movements (Shockley et al. 2003), patterns of linguistic exchange (Orsucci et al. 2006; Fusaroli and Tylén 2012), as well as more cognitive alignments.

Such a perspective connects the studies of language structure and function with both the study of embodied cognition and studies of joint action. Emergence, nature and the role of linguistic (or other symbolic, e.g., diagrammatic) communication in various task settings leads to a better understanding of the collective performance and development of common ground, through the systematic, mutual informational constraining that occurs in linguistic exchange; what happens is the selection of only certain aspects of a situation and, at the same time, stabilization of the successful symbolic patterns (Fusaroli and Tylén 2012; Galantucci

2005; Fay et al. 2010). Besides the already mentioned studies in experimental semiotics (Galantucci and Sebanz 2009), established methods such as conversational analysis are becoming particularly useful since they are designed for the study of language in its natural "habitat," in connection with the culturally specific social forms of life that take place at the moment (Schegloff et al. 1996; Zinken and Ogiermann 2011).

3.1.5 Communication Has a Coordinative Role

If we accept that symbols are selected constraints on dynamics—in other words, that they effectuate a functional reduction of the degrees of freedom of a system—in the case of natural language the coordinative function comes to the fore (see Rączaszek-Leonardi 2009a). Language lives in the interaction between and among people, thus the dynamics the symbols harness concern, above all, the dynamics of interaction. Since it is the interaction that becomes the functional system, communicative processes are thus not to be understood in terms of transmitting information, i.e., the "conduit metaphor" (Reddy 1979), in which symbols "stand for" references and thus have a power of evoking in a hearer *the same* referent that is meant by a speaker. Rather, the alternating use of symbols by participants steers the interaction as a whole through the possible state space, by constraining parts of this system in an appropriate (functional) way.

Interaction may well be the most important level at which the efficiency of symbols is "evaluated"—i.e., the effectiveness of the coordination in interaction might be the most important criterion for the selection of symbolic structures. Contrary to our phenomenal experience of mastering language as individuals, language is thus a distributed system, created by populations to exercise adaptive forms of coordination of its members. Meaning, in this picture, is co-created in participation by interlocutors (De Jaegher and Di Paolo 2007).

A striking consequence of this view is that if individual "meaning" is understood as the individual cognitive system dynamics constrained by a given symbolic structure in a given situation—i.e., the dynamical process in an individual brain evoked by a symbolic structure—then it is necessarily different for every individual. It is influenced by a personal history of the use of a given expression and an individual conceptual landscape. This challenges theories of linguistic meaning, which assume (congruently with the "conduit metaphor" of communication, the "code" metaphor for language, or the "container metaphor" of a symbol) that a symbol carries invariant meaning from a speaker to a hearer.

If, on the other hand, expressions of natural language are seen as replicable constraints on interpersonal dynamics, functional to co-action in the environment, the differences in individuals' brain dynamics that are constrained by a symbol do not pose a problem. On the contrary, for all we know, for certain coordinative purposes it might be better to have individuals with slightly different "meanings" for an item, perhaps covering a larger ground for possible joint interpretations, and/or complementing each other in a given coordinative task (Rączaszek-Leonardi and Cowley 2012).

In such populations of "differently minded" agents, the adaptation to a changing environment might be quicker than it would be in uniform, homogeneous ones, similarly to what we see in artificial classification systems using, ensemble computing, i.e., engaging agents with different algorithms.

It is true that coordinating individual physical or cognitive efforts in a situation in most cases includes agreeing on what objects in the external world the coordinated action would involve. But agreeing on reference is neither a sufficient nor a necessary condition for interpersonal coordination through language, the simplest example being the realization of the phatic function of speech (Malinowski 1981). Thus, if language is seen as a system of transmittable constraints on dynamics, it can hardly be expected that it somehow mirrors, or reflects, the external world. Any "picture" of the world's external structure detected in language will be filtered through the necessity of co-action in the world. More in the vein of Wittgenstein's language games (Wittgenstein 1953, theses 23, 241), utterances are selected as forms of life within a community (see also Goodwin 2000; Zinken 2008). They constrain existing dynamics of the participants, which includes coordination dynamics. It thus becomes crucial to identify the existing coordination in order to see how language constrains it further.

What does this mean for natural language research? Studying the harnessing (and thus enabling) role of language requires, as we said earlier, identifying the dynamics that are harnessed. This is not easy: the dynamics concern many different timescales and systems (or levels) (Rączaszek-Leonardi 2009a). One of the types of dynamics is the physical and cognitive dynamics of interaction in joint action and/or joint problem solving. This links the study of language to the studies of these phenomena. Studies of human interpersonal motor coordination have been conducted at least since the 1980s. They show powerful mechanisms for stabilizing the modes of synchronization in humans (Schmidt, Carello and Turvey 1990; Richardson et al. 2007), as well as principles for creating functional co-action systems in the face of a common task (Schmidt and Richardson 2008; Marsh et al. 2006). Linguistic interaction has its own, specific physical coordination background, with synchronized movements and imitative actions (Shockley, et al. 2003; Pickering and Garrod 2004). Neural bases for joint action are also the topic of recent research, with the findings on mirror neurons interpreted in terms of complementary rather than imitative actions (van Schie et al. 2008; Sartori et al. 2011). The study of linguistic interaction becomes a study of how the existing modes of interaction are modified by language, becoming more effective and culturally specific.

3.1.6 Degree of Constraint

Degree of constraint provides a useful dimension for characterizing expressions in terms of the role of context in determining their meaning. As we pointed out before, seeing linguistic symbols as constraints makes them naturally both context-sensitive and effective (applicable) in different contexts. Constraining, depending on the role and cultural history of an expression, may be strong (as in the case of "hydrogen peroxide"; or other scientific terms, which if given precise definitions may approach the formal languages' context-independence), weaker (as in the case of the word "dog") or very weak (as in the case of the word "here"). In natural talk, some play is always left for the natural dynamics of interaction among persons in a particular context or environment that determines the actual, realized meaning.

These features are incongruent with the picture of language as an individually represented system of arbitrary, referential symbols. Language is seen instead as i) a means of inter-individual coordination, ii) its expressions serve as doing rather than referring, and iii) the arbitrariness of linguistic symbols seems to be much more limited than usually thought. Let me elaborate on these points.

3.2 Controlled Collectivity

As already noted above, accepting the coordinating role of language changes the level that is seen as the most crucial source of selection criteria for symbolic structures. The quality of interaction in the environment, not a single action, comes to the fore. In the light of this "primacy of 'we'," the background assumed for the effective action of symbols is thus not the individual knowledge needed for understanding, but rather, the shared physical and cognitive dynamics of an already on-going interaction. The interaction, as potently shown by the work of di Paolo and his colleagues, assumes an important ontological and thus explanatory status (di Paolo et al. 2008). Certain individual traits or behaviors are simply not explainable at the level of individual cognitive processes—they appear only as a consequence of being in interaction; they are created in interaction. As pointed out in Rączaszek-Leonardi and Cowley (2012), this insight may be a step toward answering difficult questions regarding the possibility of on-line intersubjective states, joint attention, and shared goals—paraphrasing William James' view on the nature of emotions: we do not interact because we have a shared cognitive or/and emotional state, but rather we have a shared cognitive or emotional state because we interact.

This basic interactivity relies on low level, swarm-like mechanisms of interaction, which are shown, e.g., in the latest research on joint action (Richardson et al. 2010), coordination in conversation, and readiness for complementary actions (see, e.g., van Schie et al. 2008; Sartori et al. 2011). Such interactivity can be linked to distinct neural markers in individual brain activation patterns—such as the *phi*-complex (Tognoli et al. 2007). Applied to language, this shows the importance of both physical and cognitive coordination in the conceptualization of what "common ground" may consist of. When common ground is understood as shared dynamics, the richness of meaning evoked by linguistic expressions and, on the other hand, their contextuality, ceases to be a challenge. Admittedly, however, such a view does present new difficulties: (i) it makes abstract linguistic analyses difficult because one never knows what is communicated unless one knows what was shared; (ii) it imposes on a researcher the requirement to identify the many kinds of relevant dynamics. Turning again to the positive side, the view presented here may support the emerging framework for encompassing these diverse factors and point to the methods for providing data about them.

3.3 Referential Properties of Linguistic Symbols

Accepting points (d) and (e) from Section 1, in the case of human natural language (as shown in points 3.1.4 and 3.1.5) entails accentuating its coordinative role. This brings new perspectives on referential properties of language and a broader question of the possibility of cognition through language. Since elements of language in many philosophical and linguistic theories are thought to clearly correspond to objects in the world, sentences are seen as statements about these objects. Thus their meaning, casted in truth-conditional terms, corresponds to the configuration of states of the world.

Seeing language as primarily a tool for social coordination does not, obviously, eliminate the referential aspect of language. After all, coordination in most cases has to be *about* some aspects of the external world if it is to lead to the (jointly effectuated) changes in the environment. Establishing joint attention, identifying objects of joint action and joint cognition is crucial to such effectuation. However, the referential aspect is secondary to coordination and important in only some aspects of communication. What is more, reference in coordination situations, although preserving the physical identity of an object in a real physical interaction, may actually require inter-individual differences in "intension," as explained in point 3.1.5.

How the issue of reference relates to the meanings of natural language symbols that, in this framework, are defined by the effects on various kinds of dynamics, is a fascinating and complicated problem. What seems to be clear from this perspective is that the ways in which reference and joint reference are achieved in language should be studied by investigating the pragmatic aspects of language use (Hanna et al. 2003; Dale et al. 2011), rather than by studying the system by itself and assuming obvious referential properties of symbols. Investigating the integration of the constraining function, achieved through repetitive use and selection, can help uncover the referential role, which is always mediated through the coordinative role of language. Moreover, a possibility should be always kept open that the individual feeling of "grasping" a concrete, reified meaning of an expression may be subservient to the overall coordination goal of an expression, putting the "direct perception of meaning" in a broader perspective of socially interactive coordination.

From this perspective, grammatically formed sentence, seen as a statement about the world, does not serve as an atom of meaning; rather, units of meaning would be distinguished on more pragmatic grounds. On the timescale of interaction these would be acts of coordination, in which linguistic expressions play the controlling role. Motivated by coordination effectiveness, selective pressures will act on words, constructions, phrases, depending on how they enter interactive activity. Thus, in the words of Schegloff, "*the interactional matrix of grammar requires a different understanding of what should enter into a linguistic description and/or a different model of linguistic structure*" (Schegloff et al. 1996, p. 24).

3.4 Conditions on Arbitrariness

There is an obvious arbitrariness to the selection of a particular physical structure to become a constraint. But the crucial "contacts" that symbolic structures have with dynamical processes impose at least two kinds of limits: (1) the initial limits on arbitrariness: not any structure could have become a specific constraint; only the ones that actually can enter in a causal relationship with a constrained process, and those that can be reliably copied with minimal effort (an argument for this kind of non-arbitrariness would be the universality of the genetic code, as Pattee (1969) notes); and (2) historical limits on arbitrariness: since selection, acting on the level of phenotype, occurs by virtue of a particular physical structure harnessing the dynamics in a specific way, the history of these choices is crucial for the symbol structure's meaning. It cannot be arbitrarily swapped with a different structure that does not have this history.

Thus constraints, being physical structures themselves, in order to be good functional constraints harnessing the right kind of dynamics in a right way, must have the following properties: (i) they must have the potential of acting on the dynamics, (ii) they must have been selected on the basis of acting on the dynamics within a system of other constraints, and (iii) they must be reproducible in a reliable way by whatever mechanism there is to reproduce them. These factors will limit the form and the structure of the symbol system in a living organism.

The claim about arbitrariness, i.e., that the choice of a word is not determined by its meaning (referent) has been an important element of natural language theories (e.g., de Saussure 1916/1983). Hockett (1960) considered it to be one of the defining features of language, and the idea of this independence is in an intuitive way supported by the often quoted fact that names of the same things are different in deferent languages, or by the existence of polysemy and homonymy in language.[9] Obviously it was recognized, even by de Saussure himself, that arbitrariness is limited: a sign has to be phonologically plausible in a given language, and the free variation of signs is checked by historical and social factors (Chandler 1995). "As Lévi-Strauss noted, the sign is arbitrary *a priori* but ceases to be arbitrary *a posteriori* - after the sign has come into historical existence it cannot be arbitrarily changed (Lévi-Strauss 1972, 91)." Thus "(…) every sign acquires a history and connotations of its own which are familiar to members of the sign-users' culture." (Chandler 1995). The latter process is often seen as "conventionalization."

---

[9] A good introduction to the issue of arbitrariness in language can be found in Chandler (1995).

In a picture of language in which dynamics is always present, and in which symbolic structures are the outcome of selection on the basis of the effectiveness of constraining the dynamics and their transmittability, the types of limits to arbitrariness can be shown more clearly. Each one of the relevant dynamics will exert its pressures on the symbolic forms and the shape of their structures (Rączaszek-Leonardi 2009a). Forms have to be easily reproducible, with a good ratio of perceptual distinguishability to production costs (Lindblom et al. 1984); structures must be learnable, i.e., adapted to cognitive skills of a child (Deacon 1997), and, on the on-line interaction timescale structures of symbols are obviously shaped by the structure of constraints they—together—impose. Such factors, and others, act as a set of strong and weak pressures and selection biases in the cultural evolution of language (for an example of such possible biases in grammatical gender assignment, see Rączaszek-Leonardi 2010). Any particular form we encounter in a language is the outcome of these historical, selective processes. What is more, forms are "alive"—they are, in any moment, in contact with the dynamics that upholds the system and that may be responsible for changes (Steels 2006). Yet all of this is unexpected, if one looks just at the structure.

Moreover, in this picture it is easier to see how the process of conventionalization may occur. Conventionalization does not happen by an act of conventionalizing, but rather, by the spreading and the selection of forms in multiple interactions. Expressions that prove functional in an on-line interaction are stabilized (see, e.g., Fusaroli et al. 2012), and, if they sustain learnability requirements, spread as culturally-specific controls on co-action, sharpening their function in multiple social episodes (Zinken and Ogiermann 2011). Arbitrariness, obviously, is still a feature of languages, but the claim about its degree becomes much more qualified here than in the code-view of language.

I hope to have shown above, that the understanding of the notion of a "symbol" proposed here, and the view on the relation that such symbols have to the dynamics of a system within which they evolve, when applied to natural language, leads to a change of several important assumptions that many take for granted. This may result in new ways of thinking about linguistic phenomena and give a more convenient framework for studying language as a dynamical system, unfolding on several timescales and encompassing several levels of organization. This framework not only finds a place for phenomena that were not easily accounted for in previous approaches (e.g., the efficiency of language, its context dependency, coordinative nature) but also makes applicable the tools created for dealing with the said complexity: namely, the mathematical tools developed to study non-linear complex dynamics in physics. Such a perspective opens up many new dimensions on which to analyze language phenomena, such as stability, temporal coherence, dimensionality, and degree of coordination. At the same time, it is very important to note, that unlike the strictly dynamical approaches, it does not eschew the symbolic description from the explanation, nor assign it merely an auxiliary, epiphenomenal role. Symbolic structures are essential entities in this explanatory scheme, but are not understandable without their complementary dynamical description.

Claiming that the core nature of "being a replicable constraint on dynamics" is the same for biological and cultural systems brings a drastic change in the analysis of the latter, but it does not obscure the profound differences between the systems. Enumerating them is beyond the scope of this paper and, perhaps, not a sensible task in general. After all, these are very different systems that arose to constrain very different processes. Some of the important differences are described in Pattee's commentary in this volume. Others can be inferred from considering the above characteristics of symbols in biological and linguistic systems (such as emphasizing the inter-agent level of coordination in linguistic systems rather than the within-agent one). As pointed out above, considering natural language as an instance of more general informational system ushers in new dimensions on which it can be analyzed and studied. In particular, the structure of the symbolic mode is now seen as forming under multiple pressures that are specific to humans and their communication systems, thus allowing the integration of the subfields of the psychology of language in interesting and productive ways, and connecting it to other domains, such as evolution of communication, sociolinguistics or anthropology. Against this background, too, particularities of concrete systems stand out more clearly and their implications can thus be better analyzed.

3.5 Example of a Difference: Fast Replication of Un-Coded Forms

Attention to the mechanism of spreading of symbolic forms reveals the uniqueness of natural language. The non-hereditary, learning-based "replication" of linguistic structures puts the learning capacities of an individual at the center of psycholinguistic interests. Seeing these capacities as preconditions for an easy spreading of effective constraints creates a conceptual place within a bigger framework for the abilities of statistical learning, and rule formation which undoubtedly is one of the individual cognitive endowments essential for language. On the other hand, transmission of language through learning in ontogenesis is not just spreading by copying (as in the case of written language) or spreading by reproducing structure by a theoretical innate machine inside a child's head. For, aside from imitation and the statistical learning of the form, the ontogenetic processes of language learning include situated learning in which a child experiences words in action and can then try out their power in interaction. Focusing on such phenomena brings about the possibility of experimenting with, and observing the immediate effects of, the ways in which children develop a certain attitude towards language (that Cowley calls the "language stance" (Cowley 2011)), the adoption of which may drastically shorten the scale of structure selection and stabilization, as well as introduce new sources for variation and complexity.

Thus, variation and selection in biological systems proceed according to a generational clock, while human language changes on a much shorter timescale: that of cultural evolution. The individual processes through which a person, in numerous interactions, learns multiple ways of a symbol's use and has it at her disposal at the moment of production is also a source of the voluntary variability in language—i.e., innovation, which rests on the ability to use it creatively "with sense," that works alongside random variation to propose new ways of control on dynamics.

With this chapter, I hope to have shown the way in which language sciences can be transformed by changing their fundamental assumptions about the nature of language and its symbols. Some of these transformations are already taking place (as I pointed out above, mentioning the growing body of research on the dynamical factors shaping language structure as well as the research on on-line language use in real interactions), albeit without the explicit re-definition of symbols. I have merely outlined the fundamental shifts in symbol/dynamics relations, and did not explicitly address many factors crucial for the new theory of language entailed by it. Such details concern both the "dynamical side" (the question, for example, of how it happens that social and emotional coordination can be value-instilling, and what role language plays in this process (see, e.g., Trevarthen and Aitken 2001; Cowley 2004; Hodges 2007)) and the "symbolic side" which requires an in-depth examination of the "individual machinery" needed for correct and creative language use which certainly depends also on statistical learning and rule discovery.

The goal in this paper was only to show that such factors *can* be investigated within one framework—one which acknowledges that individual language "processing" is *never* separated from ongoing interpersonal and individual dynamics, and that linking to these dynamics happened through historical, selective events. These dynamics co-constitute the "meaning" of what is said and can *always* be the cause of rule breaking, creativity and flexibility. In other words, what I hoped to have shown here is that incorporating some of Howard Pattee's key insights into the study of language can result in a framework in which symbol-concentrated, structuralist analysis may find its place alongside the investigations of the dynamic forces. Giving up some long-cherished assumptions about the nature of symbols in natural language, in other words, may well be worth it. What is more, this potential for reconciliation is not limited to the natural language phenomenon, but may apply to the nature of cognition in general.

4 Further Along the Third Path of Complementarity

While Pattee has investigated the necessity of complementary models mostly at the level of genetic information in living organisms, he also recognizes it as a general principle—one that pertains to the symbol-matter or mind-body problem in general. Below I briefly sketch the history of this problem in modern cognitive science, which will constitute the background for my arguing for the value of Pattee's approach as a "third way." I will reflect on its viability towards solutions of these more general problems as well as on possible reasons why it is still not accepted by many. Finally, while agreeing that conceptually and theoretically this seems to be a promising direction, I point out that there are inherent difficulties in characterizing the brain's memory structures in terms of symbolic constraints. Two possible ways of looking at this problem will be discussed. And finally, I will end with an examination of the role of language as the external, culturally molded system of constraints on individual cognition—i.e., the workings of the mind.

4.1 "Symbolic" vs. "Dynamic" in Cognitive Science

As most cognitivists agree, the excitement of shaping the new approach to human cognition in the 1950s stemmed from, among other factors, the possibility of modeling cognitive processes in artificial intelligent systems. In many cases, this assumed the compatibility of human cognition with computational devices. Perhaps the most powerful statement of this analogy was the Turing formalism, which showed that all computation can be realized by a universal machine. As noted earlier, this led to the assumption that symbols in the human mind and computational symbols in a computer can be treated as equivalent. Quite soon, however, the theorists of cognition realized that the symbols in such formal models of thinking remain "ungrounded," and that it was not easy to specify who or what in the system could interpret the symbols—i.e., give them meaning—and how.

The most renowned critique came in the form of John Searle's Chinese Room argument, in which he addressed the insufficiency of syntactic rules for specifying semantics (Searle 1980). Similar under-specification of semantics within a formal model of mind was pointed out by Harnad (1990) in his famous paper *The Symbol Grounding Problem*. The conclusion of both works is that the interpretation of symbols, giving them meaning, requires "a loan on intelligence" (Kelso 1995), not unlike the one that behaviorists had to repay for exorcising mental states from the explanation of behavior (see Dennett 1978, who originally used the phrase).

As Pattee writes in his commentary to this volume (p. 23), Newell and Simon's Physical Symbol System Hypothesis, once proposed as the new framework for emerging cognitive sciences, did not provide any way to link the dynamics of physical laws to the non-dynamic symbol vehicles—thus, the behavior of dynamic matter was actually not thought to constitute an element of functional explanations. Cognitive systems were supposed to be realizable on many different "machines" and computations were supposed to be realized in error-free environments, usually containing no novelty.

It is important to note that what was considered to be cognitive, intelligent behavior was, in these early days of cognitive psychology and cognitive science, limited to what some would call "higher cognition," and included, above all, activities such as problem solving, conducting mathematical proofs (e.g., The General Problem Solver of Newell and Simon), reasoning, decision-making and language processing. However, after the advancement of first theories of perception and pattern recognition in this framework, and the attempts to account for so-called "lower cognition" (such as, navigating one's environment, skill acquisition, non-linguistic memory, or coordinated action) it quickly became evident that in these domains the symbol-based "information processing" was not as useful or convenient. It is primarily due to research into these range of problems that an alternative view of cognition at that time flourished, giving rise to "ecological psychology," consolidated by the research and theoretical writings of James Gibson (e.g., 1951, 1960, 1961, 1966) and subsequently continued in the Haskins Laboratories at the University of Connecticut by the Turvey group, and the Center for Complex Systems and Brain Sciences established by Scott Kelso.

The ecological psychology school stood in opposition to the information processing approach; opting for the "direct perception" of the environment, possible because of the evolutionarily "tuned-in" senses (Gibson 1966), and perception-action coupling, rather than the primacy of symbol-driven computationalism. In some forms, the approach deemed mental symbol processing unnecessary for intelligent behavior, and eschewed "symbolic mental representations," even when it came to processes that seemed to clearly require them, such as memory (Freeman and Skarda 1990). Models based on dynamics, self-organization, and local responsiveness to the environmental input could indeed account for quite an impressive range of psychological phenomena, such as speech perception and production (Kelso et al. 1984; Tuller, et al. 1994; Fowler and Saltzman 1993), motor coordination and motor learning (Kelso et al. 1990; Kelso 1995; Schöner and Kelso 1988; Schöner et al. 1992), and inter-individual coordination (Turvey 1990), at the same time becoming the foundation for novel approaches to creating artificial intelligent systems (Brooks 1991).

On the other hand, the approach was criticized for being inapplicable to the problems in "higher cognition" that most psychologists were interested in. Although attempts were made to make the notion of "symbol" compatible with the Gibsonian approach (see e.g., Greeno 1994), it was not clear what their ontological status would be and how they would be realized, incorporated into the dynamical perception-action system.

One of the first to advocate the reconciliation of the two schools was Ulric Neisser, who already in 1976 admitted his disappointment with the information-processing approach, and emphasized instead the merits of ecological psychology of perception (Neisser 1976). Later (e.g., Neisser 1994), he proposed that the way to this reconciliation would be a view of cognition as a polymorphic system, in which subsystems are responsible for dealing with specific kinds of cognitive tasks. He distinguished between direct perception, representation and recognition and the social, interpersonal sensitivity subsystems, which in higher cognitive functions come to work together.

Insightful and bold as it was, not even this approach offered a way to join symbols with dynamics; Neisser expressed a hope that the emerging connectionist approach might prove capable of integrating the models. At the end of the 1980s, with the revival of the connectionist (artificial neural networks) school, it seemed indeed that some kind of reconciliation would be possible. Neural networks were, essentially, dynamical systems, and were proposed to encompass symbolic functioning either on a sub-symbolic level (Smolensky 1988), or by associating stable "labels" represented on a set of units (e.g., Kawamoto 1993). However, it was still not clear how such "symbols" would acquire their special status (if they have one) and how, in an artificial net, their action would be different from any other input action. Some hopes arose again with the advances in hybrid modeling (ACT-R, CLARION) (McClelland 2009), which, however, were criticized for being excessively modular and most often more focused on dividing the tasks to be sent to particular modules, than in specifying the ways in which symbolic and dynamic procedures could cooperate.

The increasing popularity of such models testifies to the search for integration, which may mean that specifying the relation between symbols and dynamics in theories of behavior and cognition becomes important for a growing number of researchers. This makes one think that it is the right time for embracing an approach which long ago proposed a way for specifying this very relation.

4.2 "Symbolic" with "Dynamic" in Cognitive Science

In the light of the above, it is curious that the one approach that never decoupled symbols from dynamics was not welcomed immediately. As I suggested earlier, one possible reason was the magnetic promise of easy modeling of apparently symbolic mental processes by computational systems that strengthened the analogy and, in general, the computer metaphor of the brain. Mental "symbols" were thus taken to *be* formal symbols, i.e., symbols with "maximally harnessed dynamics." Because of this assumption, one finds the neglect of the self-reconstruction requirement of von Neumann, the sinking to near oblivion of such works as Polanyi's (1968), and the attention to Turing's ideas on the universality of computation, rather than his work on morphogenesis. Another reason for the reluctance to accept that complementary models might be necessary to deal with cognitive phenomena is the Occam's razor-like conviction that one model is always better than two, which favors reductionism (see also Pattee's commentary, p. 20). And yet another is that if material dynamics are considered important for informational processes, then one needs tools for studying and description of these dynamics—and the majority of these tools had not then been developed.

The complexity is indeed mind-boggling. As I showed earlier in this paper (and in more detail elsewhere (Rączaszek-Leonardi 2009a)), timescales of relevant dynamics for cognitive process (e.g., for natural language) range from milliseconds of brain activity, to seconds and minutes of on-line interactions, to months and years of developmental processes, to hundreds years of cultural evolution, while the mutually dependent systems that operate within these timescales include individual brains, dyads, groups and populations. Only recently have we developed mathematical tools sophisticated enough to deal with the complexity in physics and computational powers necessary for creating good models of complex phenomena. It should be pointed out that many such models in the domain of behavioral sciences were developed within the ecological psychology approach (Jeka and Kelso 1989; Schöner and Kelso 1988; Kelso 1995; Turvey 1990; van Orden et al. 2003).

With the conception that the symbolic level is necessary to describe selected constraints on dynamics, which make such symbols functional, Pattee shows that the perspectives represented by the information processing approach and the ecological psychology construct complementary models of cognitive processes, of which neither can function independently.

> I am not opposing the ecological attitude elaborated by Fowler and Turvey, nor am I supporting it in opposition to the information processing approaches. I am claiming that these are two complementary modes of description that have not yet been completely articulated and, more fundamentally, have not been recognized as essentially complementary, in the sense that any explanation of cognitive behavior will require both modes of description. Pattee (1982a, p. 22)

If the dynamics on multiple timescales is not taken into account, this means that explanation is fixated upon only a slice of dynamical process at some point of time, showing only synchronic relations; this attitude is similar to the reluctance of structural linguistics to talk about diachrony. The fact that a symbol is a physical outcome of selective processes is conveniently omitted. Stripping symbols from their history in a given dynamics entails the simplification of the system, even if it may enable a more close-up examination of their structures.

The symbolic model is thus a way of framing, generalizing, and assuming approximation of the dynamical history and agreeing that the functional framework will be external to the system. It makes sense only if one remembers such structure's dynamical provenience, which, in turn, entails that sometimes a natural symbol within its structure will not behave like a formal symbol would. The dynamic model, on the other hand, gives the account of the substrate of the symbol's controlling action—which is difficult, because often it is taken for granted. Recent rapid developments in the theory of embodiment of cognition is a sign that increasingly more attention is given to natural dynamics of the bodily interaction with environment and with others.

4.3 Complementarity of the Brain Models

According to the ideas laid out in the papers in this volume, preserving functional organization requires passing constraints from one moment in time to another. In the case of such "memories" in the brain—shall they also be seen as symbolic structures? If yes, what form could the local symbolic constraints take in this case? As Peter Cariani asks (2001, p. 61): "*What does it mean to say that neurons perform 'computations' or 'measurements' or that 'symbols' exist in the brain?*"

There are certainly many attempts to answer this question, and the summary of them is beyond the scope of this paper. In the light of the approach presented above, two ways of proceeding are perhaps most interesting: (a) looking for candidates for the replicable constraints in the brain or, (b) following, e.g., Freeman's proposal (Freeman and Skarda 1990; Freeman 1995) "leaving" the workings of the brain to the dynamical level that, being a part of a broader system, relies on externally manufactured structures for functional transmittable constraints. Let me elaborate these possibilities.

(a) Since the brain is capable of a continuous "regenerations of informational orders" (Cariani 2001, p. 63), internal constraints on dynamics must exist that serve the communication of adaptive states from one moment to another (i.e., the communication of the brain at one moment with itself at the next moment). The quest for such structures is thus open: "*The current situation in the neurosciences regarding the neural code is not unlike the situation in molecular biology before the elucidation of the genetic code*." (Cariani 2001, p. 64) Such a code, or rather, as I noted earlier, such a *language*, should explain the informational self-reconstruction and epistemic operations of the brain. Thus, the search for reproducible patterns of neuronal activity able to function as constraints.

Among the most important consequences of understanding symbolic structures as physical structures that act as constraints on dynamics is, as I pointed out earlier, that these physical structures might have a very different form in different systems (e.g., a reproducible nucleotide base or an articulatory gesture producing a soundwave in natural language) and that the constraining action is not of an all-or-none type but is a matter of a degree. In some cases (as in the case of computing devices) the harnessing is strong, to the point of not leaving any "play" to natural dynamics; in others, as in the case of the gene setting off an avalanche of dynamical events, it is just a push in the right (evolutionarily selected) direction. If the brain's memory structures can indeed be conceptualized as transmittable constraints on dynamics, then one would expect both that symbolic structures are less discrete than in the more well known symbol systems—yet still reproducible!—and that the harnessing is certainly less severe than in the case of a formal symbol. Many kinds of memory, however, can probably be explained in a Gibsonian way, as a modification of direct perception-action coupling, which leads us to the second, equally controversial, possibility:

(b) The brain does not need internal transmittable constraints, but rather relies on the external ones. Counterintuitively, the brain's ability to think symbolically would be then linked to the ability to produce external physical signs: "*Not pure ideas in pure consciousness, but concrete signs lie at the base, signs which are for us recognizable and reproducible despite small variations in detailed execution, signs which by and large we know how to handle.*" Weyl (1949), quoted in Cariani (2001, p. 70).

Within such an approach, brains are seen as parts of larger systems, together with the physical environment and other brains, with symbols constructed for social, coordinative purposes (Freeman 1995), and constituting a new cognitive niche (Clark 2006). Such a view seems congruent with Pattee's observation that: "*Until the production of the discrete symbolic expression, there is no simple concept or measure of the 'information in the brain'*." (Pattee 2006, p. 226). The general view of intelligence and cognition as intertwined with sociality is also congruent with the recently fast developing movement of distributed and extended cognition (Hutchins 1995; Clark 2008). This approach stimulates new research directions and methods in psychology but also is visible in new ways of creating artificial intelligent systems, based on multiple communicating agents (see e.g., Steels 2007; Cangelosi 2010).

In the light of the theory of natural language as a system of replicable constraints, and independently of the debate on the existence of the symbolic constraints in the brain, linguistic symbols have the power to control the dynamics of the individual brains. Since natural language arises in social interaction, these controlling processes proceed according to socially designed directions. It is through situated language, then, that the evolved forms of sociality modify individual cognition, by constraining its dynamics developmentally and on-line. From this perspective, one can gain a fresh, and perhaps more ecologically valid look at the so-called linguistic relativity question. In a sense, accepting this framework, a researcher in language becomes more Whorfian than Whorf himself: language not only designs categories for perceiving and understanding the world but goes much deeper, pervading the ways we engage with our conspecifics in coordinating action in this world (Zinken 2008).

One of the greatest consequences of Pattee's approach is that it makes us realize that the dynamics to be controlled must already be there: both the individual processes of perception/action and inter-individual coordination must be in place before language can arise and shape them further. It seems that the predominantly linguistic and symbolic way in which we express theories often makes us ignore this fact, or take it for granted. Paradoxically, the very same device that led to seeing minds as computing machines can perhaps help us in understanding the complexity of dynamics involved. Now, however—unlike in the Good Old-Fashioned Cognitive Science since the 1950s, which relied on the machines' ability to conduct formal, algorithmic processes—we will rely on the recent ability to simulate large-scale stochastic processes, which result from non-linear interactions of multiple elements, in order to reconstruct such dynamics.

References

Barbieri, M. (2003). *The organic codes – An introduction to semantic biology*. Cambridge: Cambridge University Press.

Barwise, J., & Perry, J. R. (1983). *Situations and attitudes*. Cambridge, MA: Bradford Books/The MIT Press.

Bernstein, N. (1967). *The co-ordination and regulation of movements*. Oxford: Pergamon Press.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence, 47*, 139–159.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*, 201–251.

Cangelosi, A. (2010). Grounding language in action and perception: From cognitive agents to humanoid robots. *Physics of Life Reviews, 7*(2), 139–151.

Cariani, P. (2001). Symbols and dynamics in the brain. *Biosystems, 60*, 59–83.

Chandler, D. (1995). *Semiotics for beginners*. http://www.aber.ac.uk/media/Documents/S4B/sem0a.html

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (2011). Language and other cognitive systems. What is special about language? *Language Learning and Development, 7*, 263–278.

Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Science, 10*, 370–374.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford: Oxford University Press.

Cowley, S. J. (2004). Contextualizing bodies: Human infants and distributed cognition. *Language Sciences, 26*, 565–591.

Cowley, S. J. (2011). Taking a language stance. *Ecological Psychology, 23*, 1–25.

Cowley, S. J. (2012). Cognitive dynamics: Language as values realizing activity. In A. Kravchenko (Ed.), *Cognitive dynamics and linguistic interactions* (pp. 1–32). Cambridge: Cambridge Scholars Press.

Cowley, S. J., & Love, N. (2006). Language and cognition, or, how to avoid the conduit metaphor. In A. Duszak & U. Okulska (Eds.), *Bridges and walls in metalinguistic discourse* (pp. 135–154). Frankfurt: Peter Lang.

Dale, R., Kirkham, N. Z., & Richardson, D. C. (2011). The dynamics of reference and shared visual attention. *Frontiers in Psychology, 2*, 355. doi:10.3389/fpsyg.2011.00355.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences, 6*(4), 485–507.

de Saussure, F. (1916/1983). *Course in general linguistics*. London: Duckworth.

Deacon, T. (1997). *The symbolic species*. Harmondsworth: Penguin.

Dennett, D. C. (1978). Artificial intelligence as philosophy and as psychology. In *Brainstorms*. Montgomery: Bradford Books.

Di Paolo, E. A., Rohde, M., & Iizuka, H. (2008). Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing. *New Ideas in Psychology, 26*, 278–294.

Elman, J. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science, 34*(3), 351–386.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics, 8*, 113–133.

Fowler, C. A., & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Speech Communication, 36*(2,3), 171–195.

Freeman, W. J. (1995). *Societies of brains. A study in the neuroscience of love and hate*. New York: Lawrence Erlbaum.

Freeman, W. J., & Skarda, C. A. (1990). Representations: Who needs them? In J. L. McGaugh, N. M. Weinberger, & G. Lynch (Eds.), *Brain organization and memory* (pp. 375–380). New York: Oxford University Press.

Fusaroli, R., Bahrami, B., Olsen, K., Frith, C. D., Roepstorff, A., & Tylén, K. (2012). Coming to terms: An experimental quantification of the coordinative benefits of linguistic interaction. *Psychological science, 23*(8), 931–939.

Fusaroli, R., & Tylén, K. (2012). Carving language for social coordination: A dynamical approach. *Interaction Studies, 13*(1), 103–124.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science, 29*(5), 737–767.

Galantucci, B., & Sebanz, N. (2009). Joint action: Current perspectives. *Topics in Cognitive Science, 1*, 255–259.

Galantucci, B., Kroos, C., & Rhodes, T. (2010). The effects of rapidity of fading on communication systems. *Interaction Studies, 11*, 100–111.

Gibson, J. J. (1951). What is a form? *Psychological Review, 58*, 403–412.

Gibson, J. J. (1960). The concept of the stimulus in psychology. *American Psychologist, 16*, 694–703.

Gibson, J. J. (1961). Ecological optics. *Vision Research, 1*, 253–262.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In N. O. Schiller & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 159–207). Berlin: Mouton de Gruyter.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics, 32*, 1489–1522.

Greeno, J. G. (1994). Gibson's affordances. *Psychological Review, 101*(2), 336–342.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49*, 43–61.

Harnad, S. (1990). The symbol grounding problem. *Physica D, 42*, 335–346.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98*, 74–95.

Hockett, C. F. (1960). The origin of speech. *Scientific American, 203*, 89–96.

Hodges, B. H. (2007). Good prospects: Ecological and social perspectives on conforming, creating, and caring in conversation. *Language Sciences, 29*, 584–604.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Jakobson, R. (1973/1989). *Związki językoznawstwa z innymi naukami* [Relationship between the science of language and other sciences]. Warszawa: PIW.

Jeka, J. J., & Kelso, J. A. S. (1989). The dynamic pattern approach to coordinated behavior: A tutorial review. In S. A. Wallace (Ed.), *Perspectives on the coordination of movement* (pp. 3–45). New York: North Holland.

Katz, J. J. (1966). *The philosophy of language*. New York: Harper & Row.

Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. *Language, 39*, 170–210.

Kawamoto, A. H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language, 32*, 474–516.

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: The MIT Press.

Kelso, J. A. S., DelColle, J., & Schöner, G. (1990). Action–perception as a pattern formation process. In: M. Jeannerod (Ed.), *Attention and performance XIII* (pp. 139–169). Hillsdale: Erlbaum.

Kelso, J. A. S., Tuller, B., Bateson, E.-V., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 812–832.

Kravchenko, A. V. (2007). Essential properties of language, or, why language is not a code. *Language Sciences, 29*, 650–671.

Kugler, P. N., & Turvey, M. T. (1988). Self-organization, flow fields, and information. *Human Movement Science, 7*(2–4), 97–129.

Lévi-Strauss, C. (1972). *Structural anthropology*. Harmondsworth: Penguin.

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1984). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, & O. Dahl (Eds.), *Explanations for language universals* (pp. 181–203). New York: Mouton.

Linell, P. (2005). *The written language bias in linguistics: Its nature, origins and transformations*. London: Routledge.

Lupyan, G., & Dale, R. A. (2010). Language structure is partly determined by social structure. *PLoS ONE, 5*(1), e8559. doi:10.1371/journal.pone.0008559.

MacWhinney, B. (2005). The emergence of linguistic form in time. *Connection Science, 17*(3–4), 191–211.

Malinowski, B. (1981). Problem znaczenia w językach pierwotnych. In A. Paluch (Ed.), *Malinowski*, Warszawa, pp. 258–278.

Marsh, K. L., Richardson, M. J., Baron, R. M., & Schmidt, R. C. (2006). Contrasting approaches to perceiving and acting with others. *Ecological Psychology, 18*, 1–37.

Maturana, H. R. (1978). Biology of language: The epistemology of reality. In G. A. Miller & L. Elizabeth (Eds.), *Psychology and biology of language and thought: Essays in honor of Eric Lenneberg* (pp. 27–63). New York: Academic.

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science, 1*, 11–38.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. *An account of basic findings. Psychological Review, 88*, 375–407.

Merleau-Ponty, M. (1960/1999). O fenomenologii mowy (On phenomenology of speech). In *Proza Świata: Eseje o mowie*. Warszawa: Czytelnik.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. San Francisco: WH Freeman.

Neisser, U. (1994). Multiple systems: A new approach to cognitive theory. *European Journal of Cognitive Psychology, 6*(3), 225–241.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and Search. *Communications of the ACM, 19*(3), 113–126.

Orsucci, F., Giuliani, A., Webber, C. L. Jr., Zbilut, J. P., Fonagy, P., & Mazza, M. (2006). Combinatorics and synchronization in natural semiotics, *Physica A, 361*(2), 665–676.

Oudeyer, P.-Y. (2006). *Self-organization in the evolution of speech, studies in the evolution of language*. Oxford: Oxford University Press.

Oyama, S. (2000). *The ontogeny of information* (2nd ed.). Durham: Duke University Press.

Pattee, H. H. (1968). The physical basis of coding and reliability in biological evolution. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Vol. 1, p. 67). Edinburgh: Edinburgh University Press.

Pattee, H. H. (1969). How does a molecule become a message? *Developmental Biology Supplement, 3*, 1–16.

Pattee, H. H. (1972). Laws and constraints, symbols and languages. In C. H. Waddington (Ed.), *Towards a theoretical biology 4, essays* (pp. 248–258). Edinburgh: Edinburgh University Press.

Pattee, H. H. (1973a). The physical basis and origin of hierarchical control. In H. H. Pattee (Ed.), *Hierarchy theory* (pp. 73–108). New York: G. Braziller.

Pattee, H. H. (1973b). Postscript: Unsolved problems and potential applications of hierarchy theory. In H. Pattee (Ed.), *Hierarchy theory* (pp. 131–156). New York: G. Braziller.

Pattee, H. H. (1980). Clues from molecular symbol systems. In U. Bellugi & M. Studdert-Kennedy (Eds.), *Signed and spoken language: Biological constraints on linguistic Form* (pp. 261–274).

Pattee, H. H. (1982a). The need for complementarity in models of cognitive behavior: A response to Fowler and Turvey. In W. Weimer & D. Palermo (Eds.), *Cognitive and symbolic processes* (pp. 21–30). Hillsdale: Lawrence Erlbaum Associates.

Pattee, H. H. (1982b). Cell psychology: An evolutionary approach to the symbol-matter problem. *Cognition and Brain Theory, 5*(4), 325–341.

Pattee, H. H. (1985). Universal principles of measurement and language functions in evolving systems. In C. John & K. Anders (Eds.), *Complexity language and life: Mathematical approaches* (pp. 268–281). Berlin: Springer.

Pattee, H. H. (1987). Instabilities and information in biological self-organization. In F. E. Yates (Ed.), *Self-organizing systems. The emergence of order* (pp. 325–338). New York: Plenum.

Pattee, H. H. (2006). The physics of autonomous biological information. *Biological Theory, 1*(3), 224–226.

Pattee, H. H. (2007). The necessity of biosemiotics: Matter-symbol complementarity. In M. Barbieri (Ed.), *Introduction to biosemiotics* (pp. 115–132). Dordrecht: Springer.

Pickering, M. J., & Garrod S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*, 169–225.

Polanyi, M. (1968). Life's irreducible structure. *Science, 160*, 1308–1312.

Rączaszek-Leonardi, J. (2009a). Symbols as constraints: The structuring role of dynamics and self-organization in natural language. *Pragmatics and Cognition, 17*(3), 653–676.

Rączaszek-Leonardi, J. (2009b). Metodologiczne podstawy współczesnej psycholingwistyki [Methodological bases of contemporary psycholinguistics]. In: *Metodologie językoznawstwa. Współczesne tendencje i kontrowersje.* [Current tendencies and controversies in contemporary linguistics]. Lexis: Krakow.

Rączaszek-Leonardi, J. (2010). Multiple time-scales of language dynamics: An example from psycholinguistics. *Ecological Psychology, 22*(4), 269–285.

Rączaszek-Leonardi, J., & Kelso, J. A. S. (2008). Reconciling symbolic and dynamic aspects of language: Toward a dynamic psycholinguistics. *New Ideas in Psychology, 26*, 193–207.

Rączaszek-Leonardi, J., & Cowley, S. J. (2012). The evolution of language as controlled collectivity. *Interaction Studies, 13*(1), 1–16.

Reddy, M. J. (1979). The conduit metaphor: A case of frame conflict in our language about language. In A. Ortony (Ed.), *Metaphor and thought* (pp. 284–310). Cambridge: Cambridge University Press.

Richardson, M. J., Marsh, K. L., Isenhower, R., Goodman, J., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science, 26*, 867–891.

Richardson, M. J., Marsh, K. L., & Schmidt, R. C. (2010). Challenging the egocentric view of perceiving, acting, and knowing. In L. Feldman Barrett, B. Mesquita, & E. Smith (Eds.), *The mind in context* (pp. 307–333). New York: Guilford Press.

Rosen, R. (1969). Discrete and continuous representations of metabolic models. *Quantitative biology and metabolism – Proceedings of the Helgoland 3rd International Symposium*. Springer.

Rosen, R. (1991). *Life itself: A comprehensive inquiry into the nature, origin, and fabrication of life*. New York: Columbia University Press.

Sartori, L., Cavallo, A., Bucchioni, G., & Castiello, U. (2011). From simulation to reciprocity: the case of complementary actions. *Social Neuroscience, 7*(2), 146–158.

Schegloff, E. A., Ochs, E., & Thompson, S. (1996). Introduction. In E. Ochs, E. A. Schegloff, & S. Thompson (Eds.), *Interaction and grammar*. Cambridge: Cambridge University Press.

Schmidt, R. C., & Richardson, M. J. (2008). Dynamics of interpersonal coordination. In A. Fuchs & V. Jirsa (Eds.), *Coordination: Neural, behavioural and social dynamics* (pp. 281–308). Heidelberg: Springer.

Schmidt, R. C., Carello, C., & Turvey, M. T. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 227–247.

Schöner, G., & Kelso, J. A. S. (1988). Dynamic pattern generation in behavioral and neural systems. *Science, 239*, 1513–1520.

Schöner, G., Zanone, P. G., & Kelso, J. A. S. (1992). Learning as change of coordination dynamics: Theory and experiment. *Journal of Motor Behavior, 24*(1), 29–48.

Searle, J. (1980). Minds, brains, and programs. *The Behavioral and Brain Sciences, 3*.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523–568.

Shockley, K., Santana, M.-V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of experimental psychology. Human perception and performance, 29*(2), 326–332.

Smith, K., Brighton, H., & Kirby, S. (2003). Complex systems in language evolution: The cultural emergence of compositional structure. *Advances in Complex Systems, 6*, 537–558.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11*, 1–23.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Blackwell.

Steels, L. (2006). Experiments on the emergence of human communication. *Trends in Cognitive Sciences, 10*, 347–349.

Steels, L. (2007). Fifty years of AI: From symbols to embodiment—and back. In M. Lungarella, F. Iida, J. Bongard, & R. Pfeifer (Eds.), *50 years of artificial intelligence, essays dedicated to the 50th anniversary of artificial intelligence, LNAI 4850* (pp. 18–28). Berlin: Springer.

Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences, 28*, 469–489.

Steffensen, S., & Cowley, S. J. (2010). Signifying bodies, health and non-locality: The aftermath. In S. J. Cowley, S. Steffensen, & J. C. Major (Eds.), *Signifying bodies: Biosemiosis, interaction and health* (pp. 331–355). Braga: Portuguese Catholic University Press.

Tabor, W. (1994). *Syntactic innovation: A connectionist model*. Ph.D. dissertation, Stanford University.

Thibault, P. (2004). *Brain, mind, and the signifying body: An ecosocial semiotic theory*. London/New York: Continuum.

Tognoli, E., Lagarde, J., DeGuzman, G. C., & Kelso, J. A. S. (2007). The phi complex as a neuromarker of human social coordination. *Proceedings of the National Academy of Science, 104*, 8190–8195.

Trevarthen, C., & Aitken, K. J. (2001). Infant Intersubjectivity: Research, theory and clinical application. *Journal of Child Psychology and Psychiatry, 42*(1), 3–48.

Tuller, B., Case, P., Kelso, J. A. S., & Ding, M. (1994). The nonlinear dynamics of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance, 20*(1), 3–16.

Turing, A. M. (1952). The chemical basis of morphogenesis. *Philosophical Transactions Royal Society of London B, 237*, 37–72.

Turvey, M. T. (1990). Coordination. *American Psychologist, 45*, 938–953.

van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General, 132*, 331–350.

van Schie, H. T., van Waterschoot, B. M., & Bekkering, H. (2008). Understanding action beyond imitation: Reversed compatibility effects of action observation in imitation and joint action. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1493–1500.

von Neumann, J. (1966). *Theory of self-reproducing automata.* Edited and completed by A. W. Burks. Urbana: University of Illinois Press.

Weyl, H. (1949). Wissenschaft als symbolische Konstruction des Menchens. In: Eranos Jahrbuch (pp. 427–428) as quoted in: Holton (1988). Thematic origins of scientific thought. Cambridge, MA: Harvard University Press.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell (Basil).

Wygotski, L. S. (1930/2006). *Narzędzie i znak w rozwoju dziecka*. Warsaw: Wydawnictwo Naukowe PWN. (English translation: Tool and symbol in child development). In: R. van der Veer & J. Valsiner (Eds.), *The Vygotsky Reader*. Cambridge, MA: Blackwell, (1994). Available at: http://www.marxists.org/archive/vygotsky/works/1934/tool-symbol.htm

Zinken, J. (2008). The metaphor of 'linguistic relativity'. *History and Philosophy of Psychology, 10*(2), 1–10.

Zinken, J., & Ogiermann, E. (2011). How to propose an action as objectively necessary: The case of Polish trzeba x ('one has to x'). *Research on Language and Social Interaction, 44*(3), 263–287.