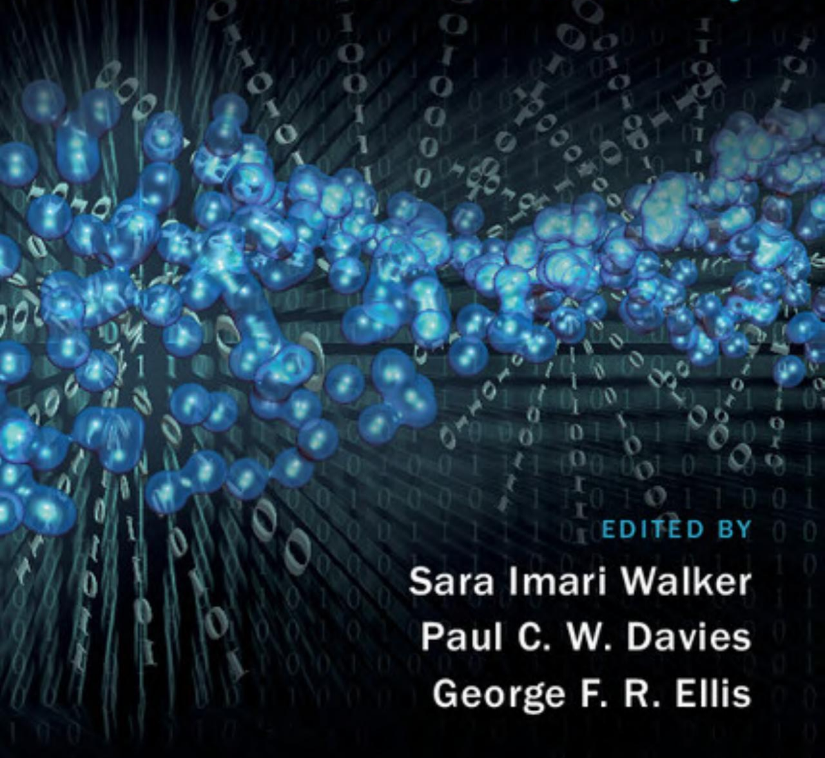
From MATTER to LIFE

Information and Causality



EDITED BY

Sara Imari Walker
Paul C. W. Davies
George F. R. Ellis

From Matter to Life

Information and Causality

Edited by

SARA IMARI WALKER
Arizona State University

PAUL C. W. DAVIES
Arizona State University

GEORGE F. R. ELLIS
University of Cape Town

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

4843/24, 2nd Floor, Ansari Road, Daryaganj, Delhi – 110002, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107150539
10.1017/9781316584200

First published 2017

Printed in the United Kingdom by TJ International Ltd. Padstow Cornwall

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging-in-Publication Data
Names: Walker, Sara Imari. | Davies, P. C. W. | Ellis, George F. R. (George Francis Rayner)
Title: From matter to life : information and causality / Sara Imari Walker, Arizona State University, Paul C. W. Davies, Arizona State University, George F. R. Ellis, University of Cape Town.
Description: New York : Cambridge University Press, 2016. | Includes bibliographical references and index.
Identifiers: LCCN 2016040374 | ISBN 9781107150539 (hardback : alk. paper)
Subjects: LCSH: Life–Origin. | Consciousness–Physiological aspects. | Biochemistry.
Classification: LCC QH325 .W35 2016 | DDC 577–dc23 LC record available at https://lccn.loc.gov/2016040374

ISBN 978-1-107-15053-9 Hardback

Contents

About the authors

CHRISTOPH ADAMI is Professor of Microbiology and Molecular Genetics, as well as Professor of Physics and Astronomy, at Michigan State University. He obtained his Ph.D. and M.A. in theoretical physics from Stony Brook University and a Diploma in Physics from Bonn University. His main research focus is Darwinian evolution, which he studies at different levels of organization (from simple molecules to brains). He has pioneered the application of methods from information theory to the study of evolution, and designed the Avida system, which launched the use of digital life as a tool for investigating basic questions in evolutionary biology. When not overly distracted by living things, he studies the fate of classical and quantum information in black holes. He wrote *Introduction to Artificial Life* (1998) and is the recipient of NASA's Exceptional Achievement Medal as well as a fellow of the American Association for the Advancement of Science (AAAS).

LARISSA ALBANTAKIS is a postdoctoral researcher with Giulio Tononi at the Department of Psychiatry at University of Wisconsin–Madison. She received her degree in physics with distinction at the Ludwig-Maximilians Universitt, Munich, followed by a Ph.D. in computational neuroscience at the Universitat Pompeu Fabra, Barcelona, under the supervision of Gustavo Deco. Her research focuses on the theoretical formulation of the integrated information theory of consciousness and its implications for evolutionary adaptation, emergence, and meaning.

ANDREW D. BRIGGS is the inaugural holder of the Chair of Nanomaterials at the University of Oxford. His research interests focus on materials and techniques for quantum technologies and their incorporation into practical devices. Current hot topics include vibrational states of nanotubes and charge transport through single molecules in graphene nanogaps. He has nearly 600 publications, with more than 16,000 citations. In 2016 his book *The Penultimate Curiosity: How Science Swims in the Slipstream of Ultimate Questions*, cowritten with Roger Wagner, was published.

PAUL C. W. DAVIES is Director for the Beyond Center for Fundamental Concepts in Science and a regents' professor at Arizona State University. He is a theoretical physicist, cosmologist, astrobiologist, and best-selling author. His research ranges from the origin of the universe to the origin of life and includes the properties of black holes, the nature of time, and quantum field theory. He is the recipient of numerous awards, including the 1995 Templeton Prize, the 2002 Michael Faraday Prize from the Royal Society, and the 2011 Robinson Prize in Cosmology.

SIMON DEDEO is an assistant professor in Complex Systems, and faculty in Cognitive Science, at Indiana University, and is an external professor of the Santa Fe Institute. At Indiana, he runs the Laboratory for Social Minds, which conducts research in cognitive science, social behavior, economics, and linguistics; recent collaborative work includes studies of institution formation in online social worlds, the emergence of hierarchy in animal conflict, competitive pricing of retail gasoline, and parliamentary speech during the French revolution. In 2017 he will join the social and decision sciences department at Carnegie Mellon University.

GEORGE F. R. ELLIS, FRS, is Professor Emeritus of Applied Mathematics at the University of Cape Town. He is a relativist and cosmologist who has worked on exact and perturbed cosmological models in general relativity theory and carefully

National Health and Medical Research Council. He spends part of each year at the University of Exeter at the Egenis, the Centre for the Study of the Life Sciences.

ANNE-MARIE GRISOGONO is a physicist by training and has worked in experimental and theoretical atomic and molecular physics, and in lasers and nonlinear optics for 14 years in various universities, followed by 20 years with the Defence Science and Technology Organisation (DSTO) working on systems design, modeling, and simulation, developing DSTO's Synthetic Environment Research Facility for defense capability development, and initiating an enabling research program into applications of complex systems science to address defense problems and future warfare concepts, for which she won a three-year Strategic Long Range Research internal fellowship. Dr. Grisogono is a member of the Australian Research Council's College of Experts, director of research for Ionic Industries, and holds an adjunct professor appointment in the engineering faculty of Flinders University. Her current research interests include fundamental questions of complexity science and improving the methodologies and tools that can be applied to dealing with complex problems.

JOSHUA A. GROCHOW is an Omidyar Fellow at the Santa Fe Institute (SFI). His overarching research dream – a distant, perhaps unreachable beacon – is a general, mathematical theory of complex systems. His primary research area is theoretical computer science, where he applies algebraic geometry and representation theory to understanding the computational complexity of algorithms. Prior to SFI he was a postdoctoral student at the University of Toronto, and he received his Ph.D. from the University of Chicago.

LUC JAEGER is a professor of chemistry and biochemistry at the University of California, Santa Barbara (UCSB). He works on informational biopolymers to further the development of new methodologies and materials with potential biomedical applications in the emerging fields of RNA nanobiotechnology and RNA synthetic biology. His research involves an effort to decipher the logic of RNA modularity and self-assembly and to unravel how complex informational molecules evolved at the origins of life. He is involved in the dialogue between science and religion and teaches a class on "What Is Life?" from both scientific and philosophical perspectives. A graduate of the University Louis Pasteur (ULP) in Strasbourg, Dr. Jaeger went on to earn a master's degree in chemistry and biology there and then a Ph.D. in structural biochemistry and biophysics at ULP in 1993 under the supervision of Eric Westhof and François Michel. He was awarded a postdoctoral research fellowship from NASA to work with Gerald Joyce at the Scripps Research Institute in La Jolla, and in 1995 returned to France as a research scientist at the Institut de Biologie Moléculaire et Cellulaire in Strasbourg. He joined the faculty of the UCSB in 2002 and was promoted to his present position in 2012. Dr. Jaeger has been a visiting professor at the National Cancer Institute and has held a grant from the ULP–National Institute of Bioscience and Human Technology and Information Services for work in Japan, as well as being the recipient of a UCSB Junior Faculty Research Incentive Award. He served as a member of the John Templeton Foundation board of advisors from 2011 to 2013. He is the author or coauthor of more than 70 papers published in scientific journals.

DAVID KRAKAUER is President and William H. Miller Professor of Complex Systems at the Santa Fe Institute (SFI). His research explores the evolution of intelligence on Earth. This includes studying the evolution of genetic, neural, linguistic, social, and cultural mechanisms supporting memory and information processing, and exploring their generalities. At each level he asks how information is acquired, stored, transmitted, robustly encoded, and processed. This work is undertaken through the use of empirically supported computational and mathematical models. He served as the founding director of the Wisconsin Institute for Discovery, the codirector of the Center for Complexity and Collective Computation, and was professor of mathematical genetics at the University of Wisconsin–Madison. He has previously served as chair of the faculty and a resident professor and external professor at SFI, a visiting fellow at the Genomics Frontiers Institute at the University of Pennsylvania, a Sage Fellow at the Sage Center for the Study of the Mind at the University of Santa Barbara, a long-term fellow of the Institute for Advanced Study in Princeton, and visiting professor of evolution at Princeton University.

THOMAS LABAR is a graduate student in the laboratory of Christoph Adami at Michigan State University and a member of the dual Ph.D. program between the department of microbiology and molecular genetics and the program in ecology, evolutionary biology, and behavior. He studies topics concerning the evolution of complexity and evolutionary innovation using digital experimental evolution. Before joining the Adami laboratory, he completed his B.Sc. degree in mathematics at Pennsylvania State University. While an undergraduate, he worked in the laboratory of Katriona Shea and researched how extinctions alter plant–pollinator communities, using computational models.

SHA LI received her B.S. degree in chemistry from Wuhan University in 2010. She then began her Ph.D. studies at Emory University working in the field of supramolecular chemistry in David Lynn's group. Her research interests include designing and engineering asymmetric peptide materials and the use of molecular self-assembly codes to develop multicomponent systems with unique electrical and optical properties. She obtained her Ph.D. degree in August 2015.

ERIC LIBBY is an Omidyar Fellow at the Santa Fe Institute where he researches the evolutionary origins of biological complexity. He is particularly interested in origins of multicellularity and the structure and shape of organismal life cycles. He completed a postdoctoral fellowship in an evolutionary biology group at the New Zealand Institute for Advanced Studies and has a Ph.D. from McGill University in quantitative physiology.

JOSEPH LIZIER is an ARC DECRA Fellow and Senior Lecturer in complex systems at the University of Sydney (since 2015). Previously he was a research scientist and postdoctoral fellow at the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Information and Communication Technology Centre (Sydney, 2012–2014), a postdoctoral researcher at the Max Planck Institute for Mathematics in the Sciences (Leipzig, 2010–2012), and has worked as a research engineer in the telecommunications industry for 10 years. He obtained a Ph.D. in computer science (2010) and bachelor's degrees in engineering (2001) and science (1999) from the University of Sydney. His research interests include information-theoretic approaches to studying dynamics in complex systems, complex networks, and neural systems, and he is a developer of the JIDT toolbox for measuring information dynamics.

DAVID G. LYNN has contributed in the general areas of molecular recognition, synthetic biology, and chemical evolution and has developed chemical and physical methods for the analysis of supramolecular self-assemblies, of signal transduction in cellular development and pathogenesis, of molecular skeletons for storing and reading information, and of the evolution of biological order. After a fellowship at Columbia University and teaching briefly at the University of Virginia and Cornell University, he served as professor of chemistry at the University of Chicago until 2000. In that year he moved to accept the Asa Griggs Candler Professorship in Chemistry and Biology at Emory University. In 2002, Lynn was awarded one of 20 inaugural Howard Hughes Medical Institute Professorships and the 2011 Emory Scholar Teacher Award for pioneering several science/arts collaborations for communicating science. He was appointed chair of the department of chemistry in 2006, as a fellow of the AAAS in 2011, and as the American Chemical Society Herty Awardee in 2013.

Chiara Marletto is a quantum physicist at the University of Oxford. After having studied Italian literature, engineering science, and condensed matter physics, she specialized in quantum computing at the University of Oxford, with a D.Phil. under the supervision of Artur Ekert. She is now collaborating with David Deutsch on the project to develop constructor theory, a new fundamental theory of physics that Deutsch proposed, which has promise for expressing in exact terms within fundamental physics emergent notions such as information, life, knowledge, and the universal constructor.

Anil K. Mehta received his Ph.D. in physical chemistry at Yale University and during this and his postdoctoral training at Washington University developed novel solid-state nuclear magnetic resonance methods for atomic-level structural characterization. He is now a faculty fellow at Emory University where he has focused on understanding the rules and forces directing molecular and supramolecular assembly and how these assemblies can be harnessed to store information and gain novel function.

Denis Noble is a physiologist and systems biologist who discovered electrical mechanisms in the proteins and cells that generate the rhythm of the heart. He constructed the first mathematical model of this process, published in *Nature* in 1960. Today, this work has grown into an international project, the Physiome Project, which constructs models of all organs of the body. He is president of the International Union of Physiological Sciences and is Professor Emeritus at Oxford University. His book, *The Music of Life*, is the first popular science book on the principles of systems biology and has been translated into many languages.

Theodore P. Pavlic is an assistant professor in the School of Computing, Informatics, and Decision Systems Engineering and

University of Sydney on the project "Postgenomic Perspectives on Human Nature."

JESPER TEGNÉR is Chaired Strategic Professor in Computational Medicine at the Centre for Molecular Medicine and Sciences for Life Laboratory at Karolinska Institutet and Karolinska University Hospital in Stockholm. He heads a research group of 35 people, one-third working in the molecular biology laboratory and two-thirds working as computational experts. He was a visiting scientist and postdoctoral fellow (Harvard) on a Wennergren three-years research position, 1998–2001, with an Alfred P. Sloan Fellowship for research. He has authored more than 100 papers, including technical computational papers as well as medical publications. He holds three undergraduate degrees (MedSchool, M.Sc.; Mathematics, M.Sc.; Philosophy), and a Ph.D./M.D. (1997) in medicine.

GIULIO TONONI is professor of psychiatry at the University of Wisconsin–Madison, where he holds the David P. White Chair in Sleep Medicine and is the Distinguished Professor in Consciousness Science. His research focuses on the understanding of consciousness and the function of sleep. His integrated information theory is a comprehensive theory of consciousness, how it can be measured, and how it is realized in the brain, which is being tested both experimentally and computationally. He is the author of more than 200 scientific publications and 5 books (2 coauthored). He has received numerous honors, including the Pioneer Award from the Director of the National Institutes of Health, and holds several patents.

SARA IMARI WALKER is an assistant professor in the School of Earth and Space Exploration and the Beyond Center for Fundamental Concepts in Science at Arizona State University and a fellow of the Arizona State University/Santa Fe Institute Center for Biosocial Complex Systems. She also serves on the board of directors of the education and research nonprofit Blue Marble Space. Her previous appointments include a NASA postdoctoral program fellowship with the NASA Astrobiology Institute and a postdoctoral fellowship in the Center for Chemical Evolution at the Georgia Institute of Technology. Her research lies at the intersection of theoretical physics, complex systems, and astrobiology and focuses on understanding the origin and nature of life.

STEVEN WEINSTEIN is a professor of philosophy at U Waterloo, cross-appointed in physics. He is also an affiliate of Perimeter Institute. Previous appointments include visiting positions at Princeton University, Dartmouth College, and the University of British Columbia. He has done pioneering work on the topic of multiple time dimensions. He also has an active career writing and recording music.

MICHAEL WIBRAL studied physics at the universities of Cologne and Konstanz, and medical physics at the University of Kaiserslautern. After working in the semiconductor industry for some years, he returned to science and obtained his Ph.D. in neurobiology at the Max Planck Institute for Brain Research in Frankfurt and the Technical University of Darmstadt. Since 2012, he is a professor for magnetoencephalography at the Goethe University, Frankfurt, and teaches information theory at the department of physics. His research interests are in neural information dynamics and predictive coding, and he is a developer of the TRENTOOL toolbox for the analysis of information transfer and active information storage in neural systems.

DAVID WOLPERT is the author of three books and more than 200 papers, has three patents, is an associate editor at more than half a dozen journals, and has received numerous awards. He has more than 10,000 citations, in fields ranging from the foundations of physics to machine learning to game theory to information theory to distributed optimization. In particular, his machine learning technique of stacking was instrumental in both winning entries for the Netflix competiton, and his papers on the no-free-lunch theorems jointly have more than 4,000 citations. He is a world expert on extending game theory to model humans operating in complex engineered systems, on exploiting machine learning to improve optimization, and on Monte Carlo methods. He is currently a member of the resident faculty of the Santa Fe Institute. Previously he was the Ulam Scholar at the Center for Nonlinear Studies, and prior to that he was at NASA Ames Research Center and was a consulting professor at Stanford University, where he formed the Collective Intelligence group. He has worked at IBM and a data-mining startup and is external faculty at numerous international institutions. His degrees in physics are from Princeton and the University of California.

HECTOR ZENIL has held positions at the Behavioural and Evolutionary Lab, Department of Computer Science, University of Sheffield; at the Structural Biology Group at the Department of Computer Science, University of Oxford; and at the Unit of Computational Medicine, Science for Life Laboratory, Centre of Molecular Medicine, Karolinska Institute in Stockholm. He is also the head of the Algorithmic Nature Group and a member of the board of directors of LABoRES in Paris. He has been visiting graduate student at the Massachusetts Institute for Technology, invited visiting scholar at Carnegie Mellon University, and invited visiting professor at the National University of Singapore. He has also been a senior research associate and external consultant for Wolfram Research, an invited member of the Foundational Questions Institute, and a member of the National Researchers System of Mexico.

1 Introduction

Sara Imari Walker, Paul C. W. Davies, and George F. R. Ellis

The concept of information has penetrated almost all areas of human inquiry, from physics, chemistry, and engineering through biology to the social sciences. And yet its status as a physical entity remains obscure. Traditionally, information has been treated as a derived or secondary concept. In physics especially, the fundamental bedrock of reality is normally vested in the material building blocks of the universe, be they particles, strings, or fields. Because bits of information are always instantiated in material degrees of freedom, the properties of information could, it seems, always be reduced to those of the material substrate. Nevertheless, over several decades there have been attempts to invert this interdependence and root reality in information rather than matter. This contrarian perspective is most famously associated with the name of John Archibald Wheeler, who encapsulated his proposal in the pithy dictum 'it from bit?' (Wheeler, 1999).

In a practical, everyday sense, information is often treated as a primary entity, as a 'thing in its own right' with a measure of autonomy; indeed, it is bought and sold as a commodity alongside gas and steel. In the life sciences, informational narratives are indispensable: biologists talk about the genetic code, about translation and transcription, about chemical signals and sensory data processing, all of which treat information as the currency of activity, the 'oil' that makes the 'biological wheels go round'. The burgeoning fields of genomic and metagenomic sequencing and bioinformatics are based on the notion that informational bits are literally vital. But beneath this familiar practicality lies a stark paradox. If information makes a difference in the physical world, which it surely does, then should we not attribute to it causal powers? However, in physics causation is invariably understood at the level of particle and field interactions, not in the realm of abstract bits (or qubits, their quantum counterparts). Can we have both? Can two causal chains coexist compatibly? Are the twin narratives of material causation and informational causation comfortable bedfellows? If so, what are the laws and principles governing informational dynamics to place alongside the laws of material dynamics? If not, and information is merely an epiphenomenon surfing on the underlying physical degrees of freedom, can we determine under what circumstances it will mimic autonomous agency? This volume of essays addresses just such foundational questions. It emerged from a 2014 workshop on 'Information, Causality and the Origin of Life' hosted by the Beyond Center for Fundamental Concepts in Science at Arizona State University as part of a Physics of Living Matter series and funded by the Templeton World Charity Foundation under their 'Power of Information' research programme. Contributors included physicists, biologists, neuroscientists, and engineers. The participants were tasked with addressing the question: Is information merely a useful explanatory concept or does it actually have causal power? Among the questions considered were:

- What is information? Is it a mere epiphenomenon or does it have causal purchase?
- How does it relate to top-down causation?
- Where does it come from?
- Is it conserved, and if so, when?
- How and when does it take on 'a life of its own' and become 'a thing' in its own right?
- Are there laws of 'information dynamics' analogous to the laws of material dynamics? Are they the same laws? Or can information transcend the laws of mechanics?
- How does information on the microscale relate to information on the mesoscale and macroscale (if at all)?
- Is information loss always the same as entropy?

Although participants agreed that the concept of information is central to a meaningful description of biological processes, opinions differed over whether the ultimate explanation for life itself necessarily rests on informational foundations. Related to this dichotomy argue that a complete mechanistic account of living matter will still fail to capture what it means to be living. By implication, some new physical laws or principles will be needed. Although these laws and principles are not known at this time, Walker and Davies argue that a promising area to seek them is precisely in the realm of information theory and macrostates, specifically that macrostates have causal power. To elaborate on this quest, their chapter discusses the issue of what exists and why the world is complex in a manner that a causal theory of information could potentially explain. Tackling the hard problems of both life and consciousness within the framework of information theory and causal counterfactuals holds the promise of an eventual unification of biology and neuroscience at a fundamental conceptual level.

Another contribution along these general lines is Chapter 3, by Marletto, in which she rejects the standard formulation of information being a facet of probability distributions or inferential knowledge. Instead, she appeals to constructor theory to argue that 'information' implicitly refers to certain interactions being possible in nature (e.g., copying) and that the properties we associate with information are constraints on the laws of physics themselves. Constructor theory is an entirely new mode of explanation, which attempts to reformulate all of science by introducing information as a foundational concept. Thus, one can discuss what is necessary of the laws of physics in order for processes such as self-replication and adaptation – central to life – to exist. Marletto considers in particular how the principles of constructor theory can account for the very existence of life (defined as accurate self-reproduction) under 'no-design laws' (Wigner, 1961) – that is, laws of physics that do not explicitly include the design of an organism at the outset. Marletto concludes that accurate constructors are indeed permitted under no-design laws, provided that the laws of physics allow the physical instantiation of modular, discrete replicators, which can encode algorithms or 'recipes' for the construction of the replicator. A novelty of the foregoing approaches is that they invert the usual assumption that physics informs biology. Since life is obviously permitted by the laws of physics, we can ask how the existence of life can inform our understanding of physical law.

While it is crucial to understand how the existence of life is consistent with (possibly new) laws of physics, establishing that fact would still leave open the question of *how life emerges from nonlife*. (To say that B is consistent with A does not imply that A explains B.) It is an open question of whether *bio emerged from bit*, that is, whether a better understanding of the concepts of information and causation have anything to say about the transition from the nonliving to living state. In Chapter 4, Grisogono asks whether there was a time before information itself emerged. While there is an elementary sense in which information existed prior to life, in terms of the Shannon information describing the configurations of matter (which Grisogono refers to as 'latent' information), the nature of biological information goes beyond the mere syntactics of Shannon. The 'added value' implicit in biology is semantics and is eloquently captured by Bateson's famous dictum (1972):

> What we mean by information – the elementary unit of information – is a difference that makes a difference.

The physical aspects of information do not touch the key issue that all living systems have purpose or function (Hartwell et al., 1999), and that purpose is realised by use of information deployed so as to enable the organism to attain its goals, enabled by architectures that encode, decode, and utilise information to sense the environment and respond in a suitable way. Grisogono addresses this distinction by pointing out three significant features of information: (1) it differs qualitatively from matter or energy, (2) it can have causal effects, and (3) not all differences make a difference. It is in the emergence of 'differences that make a difference', she argues, that the key to understanding the origin of life lies. A scenario is envisaged where the steps towards life are initiated with the emergence of an autocatalytic set of molecules that can collectively reproduce and create as a by-product membrane molecules that enclose the set (e.g. form a boundary), a set of ideas proposed in Deacon's autocell model (Deacon, 2006). While these structures would certainly contain information, the key transition that makes a difference is the origin of meaning – defined as when an autonomous agent emerges that can make an informed rather than random choice between at least two actions (counterfactuals), i.e., when it can take action on differences that make a difference. This chapter therefore connects the emergence of meaning (semantic information) and *coded* information to the emergence of causally effective information (Walker and Davies, 2013), a critical step in the origin of life.

Although a major thrust in this volume is to understand information as a distinct category, it is agreed that it cannot be 'free-floating'. As proclaimed by Landauer, 'information is physical!' (Landauer, 1996), by which he meant that every bit of information must be instantiated in a physical degree of freedom. Physical and chemical constraints on how information is processed therefore do matter. In particular, life employs both digital and analogue information, both of which may have emerged early in the transition from matter to life. In Chapter 5, Smith-Carpenter et al. consider chemistries that could permit emerging codes through the processes of chemical evolution, using self-assembling peptide β-sheets as an explicit example of a more general theoretical framework. They identify three necessary dimensions: (1) structural order, (2) dynamic networks, and (3) function. It is at the intersection of these dimensions that pathways to increasing complexity are possible, they argue, suggesting new modes for thinking about chemical evolution that are neither strictly digital nor analogue. An interesting feature of the peptide assemblies discussed is that there exists both a digital and an analogue aspect to their information content: digital in the sequence of amino acids composing the peptide, but analogue in the conformations that macroassemblies can assume.

The digital nature of life's information processing systems is familiar – DNA, for example, is best described as a digital (discrete) sequence of nucleobases, and gene regulatory networks are often described using the operations of binary logic (Davidson and Levin, 2005), such that they may be likened to circuit boards (Nurse, 2008). Less appreciated are analogue aspects of information as it operates in organisms. These are explored in Chapter 6 by Noble, who asks: 'Are organisms encoded as purely digital molecular descriptions in their gene sequences?' The answer he provides is a resounding No! One argument in favour of analogue information as a major contributor to biological function is the sheer magnitude of the information encoded in structural versus genomic degrees of freedom within a cell: a back-of-the-envelope calculation reveals that it is easy to represent the three-dimensional image structure of a cell as containing much more information than is possible in the genome. This perspective is parsimonious if one considers the evolutionary advantage of *not* encoding everything the cell does – why code for what physics and chemistry can do for you? Noble invokes a computer analogy, suggesting that not only do we need the 'program' of life; we also need the 'computer' of life (the interpreter of the genome), i.e. the highly complex egg cell, to have a full explanatory account of information in living systems.

The analogy between biology and computation brings to light the question of how much of life can be understood in terms of informational software or programs (be they genetic or in other forms) that transcend the chemical substrates of life. In Chapter 7, Adami and LaBar take one extreme, considering a purely informational definition for life as 'information that copies itself' and explore the consequences for our understanding of the emergence of life, utilising the digital life platform Avida as a case study. Based on information-theoretic considerations, they demonstrate that it is rare, but not impossible, to find a self-replicating computer program purely by chance. However, the probability is significantly improved if the resource distribution is biased in favour of the resource compositions of self-replicators, that is, if one uses a 'biased typewriter'. The conclusion is that the composition of the prebiotic milieu really does matter in determining how likely it is to stumble across a functional replicator by chance.

Flipping this narrative on its head, it suggests new information-theoretic approaches to determining the optimal environments from which life is *most probable* to emerge.

New approaches to information are necessary for understanding not only the origin of life but also all levels of description within the biological (information) hierarchy – from cells to societies. It is likely the case that insights from other (higher) scales of organisation in living systems will ultimately inform our understanding of life's emergence, and in particular that new principles will be necessary to unify our understanding of life as it exists across different scales of space and time. One such 'hidden principle' proposed by Krakauer in Chapter 8 suggests that life can be thought of as a collection of evolved inferential mechanisms dependent on both information (memory storage) and information dynamics – e.g., 'computation'. Computation provides an advantage for adaptive search in the quest for more efficient means to utilise available free energy gradients. Many examples of what he calls cryptosystems are provided, ranging from parasites such as *Trypanosoma brucei*, which uses 'noise' to evade its host's immune system by combinatorially rearranging its surface proteins, to combinatorial ciliates, which hide information by encrypting their own genome. It seems that once one starts looking for encrypted informational systems in biology, they are everywhere.

It is a surprising twist that noise could find a constructive role in biology by encrypting valuable information; noise is usually regarded as the antithesis of information in our standard physical interpretations. Another area where noise takes on a surprising role in the informational narrative of living systems is with respect to function. In nonlinear systems introducing noise can preferentially amplify functional aspects of a signal above a threshold for detection, as occurs in the phenomenon of *stochastic resonance*, a concept derived from engineering due to its utility in increasing the signal-to-noise ratio of a signal. In Chapter 9, Weinstein and Pavlic explore how biological systems might equivalently utilise noise to execute biological function. Examples are provided, such as nest-site selection meaningless in a specific context. This process, leading to the storing of genetic information in the DNA of cells, underlies the emergence of complex life. One can make the case that each of the major inventions of evolutionary history was the discovery in this way of new means of deploying information to control material outcomes in such a way as to enhance survival prospects. Farnsworth and colleagues argue that top-down causal control and the resulting appearance of autonomy are hallmarks of life.

Adopting a framework where one explicitly focuses on causal structure – rather than dynamics through state space – may help to elucidate some of the debate. This is the perspective provided in Chapter 14 by Albantakis and Tononi, who consider the distinction between 'being' and 'happening', utilising cellular automata (CA) as a case study. Most prior work on dynamical systems, including CA, focuses on what is 'happening' – the dynamical trajectory of the system through its state space – that is, they take an extrinsic perspective on what is observed. Often, complexity is characterised using statistical methods and information theory. In a shift of focus to that of causal architecture, Albantakis and Tononi consider what the system 'is' from its own *intrinsic* perspective, utilising the machinery of integrated information theory (IIT), and demonstrate that intrinsic (causal) complexity (as quantified by integrated information Φ in IIT) correlates well with dynamical (statistical) complexity in the examples discussed. These and similar approaches could provide a path forward for a deeper understanding of the connection between causation and information as hinted at in the beginning of this chapter.

Further insights into unifications of the concept of information and causation are provided in Chapter 15 by Stotz and Griffiths, who focus on the concept of biological specificity to illuminate the relationship between biological information and causation. They propose that causal relationships in biology are 'informational' relationships simply when they are highly specific, and introduce the idea of 'causal specificity', adopted from the literature on the philosophy of causation, as a way to quantify it. An example is 'Crick information', defined such that if a 'cause' makes a specific difference to the linear sequence of a biomolecule, it contains Crick information for that molecule (Griffiths and Stotz, 2013). Nucleic acids are one example of causal specificity. However, the general theoretical framework is expected to apply to a wide array of biological systems where causal specificity plays an important role.

Another example concerns animal nervous systems, which are hard-wired to collect information about the world through multiple sensory modalities. Brains are exquisitely structured to search for patterns in that information in the light of the current context and expected future events, so as to extract what is meaningful and to ignore the rest. Language has been developed to store, analyse, and share that information with others, thus permitting the transfer of cultural information. The ability of humans to transmit information qualitatively distinguishes them from the rest of the animal kingdom. Hence information acquisition, analysis, and sharing through the use of language are core aspects of what it means to be human. The information is *specifically* encoded via symbolic systems such as writing. Thus, many aspects of the role of information and causation in life and its origins in the preceding discussion are apparent in social and technological systems. Despite more than 3.5 billion years obscuring reconstruction of the events surrounding life's origin(s), these connections suggest that perhaps common principles might underlie the transition from matter to life and the current transitions we are undergoing in human social and technological systems.

In Chapter 16, DeDeo details major transitions in the structure of social and political systems, drawing insights from major evolutionary transitions more broadly. He identifies three critical stages in the emergence of societies, each of which relies on the relationship of human minds to coarse-grained information about their world. The first is awareness and use of group-level social facts (e.g., a social hierarchy as provided by the example of the monk parakeet). In a second transition, these facts eventually become norms, forming a notion of the way things 'should be' (i.e., such as thanking a shopkeeper after visiting their shop). In the third transition, the norms aggregate into normative bundles, which establish group-level relationships among norms. It is intriguing to speculate that the role of top-down causation in the origin of society as outlined by DeDeo could parallel that conjectured to occur in the emergence and evolution of life (Ellis, 2011; Noble, 2012; Walker and Davies, 2013), particularly through major evolutionary transitions (Maynard Smith and Szathmary, 1997; Walker et al., 2012).

Among all life's information processing, none evokes more appreciation for the causal power of information than that of the human mind. Explaining behaviour in terms of information processing has been a fundamental commitment of cognitive science for more than 50 years, leading to the huge strides made in psychology, linguistics, and cognitive neuroscience. Although the success of these sciences puts the reality of neural information processing beyond serious doubt, the nature of neural information remains an unanswered foundational question. This is a topic addressed in Chapter 17 by Wibral et al., who discuss how implementing techniques from information theory can aid in identification of the algorithms run by neural systems, and the information they represent. A computational principle adopted by many neural information-processing systems is to continuously predict the likely informational input and to carry out information processing primarily on error signals that record the difference between prediction and reality. This predictive coding actually shapes the way we experience the world (Frith, 2013). Wibral et al. provide insights for identifying the algorithms that are applicable to both natural and artificial systems – perhaps inspiring the design of artificial systems.

The implications of these kinds of approaches are potentially profound. A foundational, deep understanding of living systems, of the kind we currently enjoy in other domains of science such as quantum theory or general relativity, would undoubtedly dramatically change our perceptions of the world and our place in it, just as was the case for previous scientific revolutions. A hint at the implications of a fundamental understanding of the role of information in life and mind is the topic of discussion in Chapter 18 by Briggs and Potgeiter (Chapter 18), who consider the specific example of the scientific and ethical implications of machine learning. They ask: To what extent can the mechanisms that underlie machine learning mimic the mechanisms involved in human learning? An important distinction between the human mind and any machine we have yet created is that while the human brain can learn algorithms, its natural *modus operandi* is holistic pattern recognition based in neural networks (the information is integrated [Kandel, 2012]). The laws of physics themselves also do not appear to be algorithmic in nature (but instead appear to be descriptions of interactions between fields/forces and particles with fixed dynamical laws). It is therefore unclear at present how far our current approaches can take us in realising 'machines that think' (or feel) without addressing both the hard problems of consciousness and life.

REFERENCES

Bateson, G. (1972). *Steps to an ecology of mind: collected essays in anthropology, psychiatry, evolution, and epistemology*. University of Chicago Press.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, **2**(3):200–219.

Davidson, E., and Levin, M. (2005). Gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(14):4935.

Davies, P. C. W., and Lineweaver, C. H. (2011). Cancer tumors as metazoa 1.0: tapping genes of ancient ancestors. *Physical Biology*, **8**(1):015001.

Deacon, T. W., (2006). Reciprocal linkage between self-organizing processes is sufficient for self-reproduction and evolvability. *Biological Theory*, **1**(2):136–149.

Deacon, T., and Sherman, J. (2008). The pattern which connects pleroma to creatura: the autocell bridge from physics to life. In Jesper Hoffmeyer (ed), *A legacy for living systems: Gregory Bateson as precursor to biosemiotics*, pages 59–76. Springer.

Deutsch, D. (2013). Constructor theory. *Synthese*, **190**(18):4331–4359.

Donaldson-Matasci, M. C., Bergstrom, C. T., and Lachmann, M. (2010). The fitness value of information. *Oikos*, **119**(2):219–230.

Ellis, G. F. R. (2011). Top-down causation and emergence: some comments on mechanisms. *Interface Focus*, **2**:126–140.

Flack, J., Erwin, D., Elliot, T., and Krakauer, D. (2013). Timescales, symmetry, and uncertainty reduction in the origins of hierarchy in biological systems. In Kim Sterelny, Richard Joyce, Brett Calcott, and Ben Fraser (eds), *Cooperation and its evolution*, pages 45–74, MIT Press.

Frith, C. (2013). *Making up the mind: how the brain creates our mental world*. John Wiley & Sons.

Griffiths, P., and Stotz, K. (2013). *Genetics and philosophy: an introduction*. Cambridge University Press.

Hartwell, L. H., Hopfield, J. J., Leibler, S., and Murray, A. W. (1999). From molecular to modular cell biology. *Nature*, **402**:C47–C52.

Kandel, E. (2012). *The age of insight: the quest to understand the unconscious in art, mind, and brain, from Vienna 1900 to the present*. Random House.

Küppers, B. (1990). *Information and the origin of life*. MIT Press.

Landauer, R. (1996). The physical nature of information. *Physics Letters A*, **217**(4):188–193.

Maynard Smith, J., and Szathmary, E. (1997). *The major transitions in evolution*. Oxford University Press.

Noble, D. (2012). A theory of biological relativity: no privileged level of causation. *Interface Focus*, **2**(1):55–64.

Nurse, P. (2008). Life, logic and information. *Nature*, **454**(7203):424–426.

Shalizi, C., and Moore, C. (2003). What is a macrostate? Subjective observations and objective dynamics. arXiv preprint cond-mat/0303625.

Shannon, C. (1949). Communication theory of secrecy systems. *Bell System Technical Journal*, **28**(4):656–715.

Walker, S., and Davies, P. (2013). The algorithmic origins of life. *J. R. Soc. Interface*, **10**(79):20120869.

Walker, S. I., Cisneros, L., and Davies, P. C. W. (2012). Evolutionary transitions and top-down causation. arXiv preprint arXiv:1207.4808.

Wheeler, J. A. (1999). Information, physics, quantum: the search for links. In A. J. G. Hey (ed), *Feynman and computation*. Perseus Books.

Wigner, E. P. (1961). The probability of the existence of a self-reproducing unit. In *The logic of personal knowledge: essays presented to M. Polanyi on his seventieth birthday, 11th March, 1961*, pages 231–238.

Yockey, H. P. (2005). *Information theory, evolution, and the origin of life*. Cambridge University Press.

2 The "Hard Problem" of Life

Sara Imari Walker and Paul C. W. Davies

There are few open problems in science as perplexing as the nature of life and consciousness. At present, we do not have many scientific windows into either. In the case of consciousness, it seems evident that certain aspects will ultimately defy reductionist explanation, the most important being the phenomenon of qualia – roughly speaking, our subjective experience as observers. It is a priori far from obvious why we should have experiences such as the sensation of the smell of coffee or the blueness of the sky. Subjective experience isn't necessary for the evolution of intelligence (we could, for example, be zombies in the philosophical sense and appear to function just as well from the *outside* with nothing going on *inside*). Even if we do succeed in eventually uncovering a complete mechanistic understanding of the wiring and firing of every neuron in the brain, it might tell us nothing about thoughts, feelings, and what it is like to experience something. Our phenomenal experiences are the only aspect of consciousness that appears as though they cannot, *even in principle*, be reduced to known physical principles. This led Chalmers to identify pinpointing an explanation for our subjective experience as the "hard problem of consciousness." The corresponding "easy problems" (in practice not so easy) are associated with mapping the neural correlates of various experiences. By focusing attention on the problem of subjective experience, Chalmers highlighted the truly inexplicable aspect of consciousness, based on our current understanding. The issue, however, is by no means confined to philosophy. Chalmers' proposed resolution is to regard subjective consciousness as an irreducible, fundamental property of mind, with its own laws and principles. Progress can be expected to be made by focusing on what would be required for a theory of consciousness to stand alongside our theories for matter, even if it turns out that something fundamentally new is not necessary.

The same may be true for life. With the case of life, it seems as though we have a better chance of understanding it as a physical phenomenon than we do with consciousness. It may be the case that new principles and laws will turn out to be unnecessary to explain life, but meanwhile their pursuit may bring new insights to the problem (Cronin and Walker, 2016). Some basic aspects of terrestrial biology – for example, replication, metabolism, and compartmentalization – can almost certainly be adequately explained in terms of known principles of physics and chemistry, and so we deem explanations for these features to belong to the "easy problem" of life. Research on life's origin for the past century, since the time of Oparin and Haldane and the "prebiotic soup" hypothesis, has focused on the easy problem, albeit with limited progress. The more pressing question, of course, is whether all properties of life can in principle be brought under the "easy" category, and accounted for in terms of known physics and chemistry, or whether certain aspects of living matter will require something fundamentally new. This is especially critical in astrobiology; without an understanding of what is meant by "life" we can have little hope of solving the problem of its origin or to provide a general-purpose set of criteria for identifying it on other worlds. As a first step in addressing this issue, we need to clarify what is meant by the "hard problem" of life, that is, to identify which aspects of biology are likely to prove refractory in attempts to reduce them to known physics and chemistry, in the same way that Chalmers identified qualia as central to the hard problem of consciousness. To that end we propose that *the hard problem of life is the problem of how "information" can affect the world*. In this chapter we explore both why the problem of information is central to explaining life and why it is hard, that is, why we suspect that a full resolution of the hard problem will not ultimately be reducible to known physical principles (Walker, 2015).

WHY THIS "HARD PROBLEM"?

There is an important distinction between the hard problem of life and that of consciousness. With consciousness it is obvious to each of us that we experience the world – to read this page of text you are *experiencing* a series of mental states, perhaps a voice reading aloud in your head or a series of visual images. The universal aspects of experience are therefore automatically understood to each of us: if intelligent aliens exist and are also conscious observers like us, we might expect the objective fact that they experience the world to be similar (even if the experience itself is subjectively different), despite the fact that we can't yet explain what consciousness is or why it arises. By contrast, there is no general agreement on what features of life are universal. Indeed, they could be so abstract that we have yet to identify them (Davies and Walker, 2016).

As Monod (1974) emphasized, biological features are a combination of chance and necessity, combining both frozen accidents and law-like evolutionary convergence. As a result of our anthropocentric vantage point (Carter and McCrea, 1983) (thus far observing life only here on Earth), both astrobiology and our assumptions about nonhuman consciousness tend to be biased by our understanding of terrestrial life. With only one sample of life at our disposal, it is hard to separate which features are merely accidental, or incidental, from the "law-like" features that we expect would be common to *all life* in the universe.

Discussions about universal features of life typically focus on chemistry. In order to generalize "life as we know it" to potential universal signatures of life, we propose to go beyond this emphasis on a single level (chemistry) and recognize that *life might not be a level-specific phenomenon* (Walker et al., 2012). Life on Earth is characterized by hierarchical organization, ranging from the level of cells to multicellular organisms to eusocial and linguistic societies (Szathmary and Maynard Smith, 1994). A broader concept of life, and perhaps one that is therefore more likely to be universal, could be applied to multiple levels of organization in the biosphere – from cells to societies – and might in turn also be able to describe alien life with radically different chemistries. The challenge is to find universal principles that might equally well describe any level of organization in the biosphere (and ones yet to emerge, such as speculated transitions in social and technological systems that humanity is currently witnessing, or may one day soon witness). Much work has been done attempting to unify different levels of organization in biological hierarchies (see, e.g., Campbell, 1974; Szathmary and Maynard Smith, 1994), and although we do not yet have a unified theory, many authors have pointed to the concept of information as one that holds promise for uncovering currently hidden universal principles of biology at any scale of complexity (e.g., Davies and Walker, 2016; Farnsworth et al., 2013; Flack et al., 2013; Jablonka and Szathmáry, 1995; Smith, 2008; Szathmary and Maynard Smith, 1994; Walker et al., 2012; to name but a few) – ones that in principle could be extrapolated to life on other worlds.

Although we do not attempt in this chapter to define "biological information," which is a subject of intense debate in its own right (Godfrey-Smith and Sterelny, 2008), we wish to stress that it is not a passive attribute of biological systems, but plays an active role in the execution of biological function (see, e.g., Chapter 15). An example from genomics is an experiment performed at the Craig Venter Institute, where the genome from one species was transplanted to another and "booted up" to convert the host species to the foreign DNA's phenotype – quite literally reprogramming one species into another (Lartigue et al., 2007). Here it seems clear that it is the *information* content of the genome – the sequence of bits – and not the chemical nature of DNA as such that is (at least in part) "calling the shots." Of course, a hard-nosed reductionist might argue that, in principle, there must exist a purely material narrative of this transformation, cast entirely in terms of microstates (e.g., events at the molecular level). However, one might describe this position as "promissory reductionism," because there is no realistic prospect of ever attaining such a complete material narrative or of its being any use in achieving an understanding of the process even if it were attained. On practical grounds alone, we need to remain open to the possibility that the causal efficacy of information may amount to more than a mere methodological convenience and might represent a new causal category not captured in a microstate description of the system. What we term the "hard problem of life" is the identification of the actual physical mechanism that permits information to gain causal purchase over matter. This view is not accommodated in our current approaches to physics.

WHAT IS POSSIBLE UNDER THE KNOWN LAWS OF PHYSICS?

Living and conscious systems attract our attention because they are highly remarkable and very special states of matter. In the words of the Harvard chemist George Whitesides,

> How remarkable is life? The answer is: very. Those of us who deal in networks of chemical reactions know of nothing like it. How could a chemical sludge become a rose, even with billions of years to try?
>
> *(Whitesides, 2012)*

The emergence of life and mind from nonliving chemical systems remains one of the great outstanding problems of science. Whatever the specific (and no doubt convoluted) details of this pathway, we can agree that it represents a transition from the mundane to the extraordinary.

In our current approaches to physics, where the physical laws are fixed, any explanation we have for why the world is such as it is ultimately boils down to specifying the initial state of the universe. Since the time of Newton, our most fundamental theories in physics have been cast in a mathematical framework based on specifying an initial state and a deterministic dynamical law. Under this framework, while physical states are generally time dependent and contingent, the laws of physics are regarded as timeless, immutable, standard picture, we require special initial conditions to explain the complexity of the world, but also have a sense that we should not be on a particularly special trajectory to get here (or anywhere else), as it would be a sign of fine-tuning of the initial conditions. Stated most simply, a potential problem with the way we currently formulate physics is that you can't necessarily get everywhere from anywhere (see Walker, 2016 for discussion).

A real living system is neither deterministic nor closed, so an attempt to attribute life and mind to special initial conditions would necessarily involve fixing the entire cosmological initial state to arbitrarily high precision, even supposing it were classical. If instead one were to adopt a quantum cosmological view, then the said pathway from the mundane to the extraordinary could, of course, be accommodated within the infinite number of branches of the wave function, but again this is scarcely a scientific explanation, for it merely says that anything that can happen, however extraordinary, will happen somewhere within the limitless array of histories enfolded in the wave function.

Leaving aside appeal to special initial conditions, or exceedingly unusual branches of cosmological wave functions, one may still ask whether pathways from the mundane to the extraordinary are problematic within the framework of known physics. Here we wish to point out a less appreciated fact with respect to the problem of fine-tuning and explaining the complexity of our world. Just because every intermediate state on a pathway to a novel state is physically possible does not mean that an arbitrary succession of states is *also possible*. If we envisage the route from mundane chemistry to life, and onward to mind, as a trajectory in some enormous state space, then not every trajectory is consistent with the known laws of physics. In fact, it may well be that almost all trajectories are inconsistent with the known laws of physics (this could be true even if individual steps taken along the way are compatible with known laws).

To justify this claim we explore a toy model inspired by cellular automata (CA), which are often used as computational models for exploring aspects of living and complex systems. However, we note that our arguments, as presented here, are by no means exclusive to CA and could apply to any discrete dynamical system with fixed rules for its time evolution. We note that it is not necessarily the case that the physical laws governing our universe are completely deterministic (for example, under collapse interpretations of quantum theory) and that reality is not necessarily discrete. However, by demonstrating a proof-of-principle for the more conservative case of discrete deterministic systems we expect that at least some aspects will be sufficiently general to apply to physical laws, as they might describe the real universe under assumptions more relaxed than those presented herein.

CA are examples of discrete dynamical systems that consist of a regular grid of cells, each of which can be in a finite number of states – in particular, we focus on systems with cells that can be in one of two possible states: "0" or "1." For simplicity, let's also assume our universe is one-dimensional with a spatial size of w cells. The configuration space of the system then contains 2^w possible states. If we restrict ourselves to deterministic systems, saying nothing yet about the laws that operate on them, each state may appear exactly once on any given trajectory, prior to settling into an attractor (otherwise, the system would not be deterministic). Under this constraint, the total number of deterministic trajectories of length $r \leq 2^w$, $n_t(r)$, is just the number of possible permutations on a subset r chosen from 2^w elements:

$$n_t(r) = \frac{2^w!}{(2^w - r)!} \tag{2.1}$$

which quantifies the number of ways to obtain an *ordered* subset of r elements from a set of 2^w elements. The total number of unique possible trajectories is just the sum over all possible trajectory lengths r:

$$N = \sum_{r=1}^{2^w} \frac{2^w!}{(2^w - r)!} = e\Gamma(1 + 2^w, 1) - 1 \tag{2.2}$$

where $\Gamma(x, a)$ is the incomplete gamma function. The above includes enumeration over all nonrepeating trajectories of length 2^w and trajectories of shorter length that settle to an attractor at a time $r < 2^w$. Here, *N should be interpreted as the number of total possible deterministic trajectories through a configuration space, where states in the space each contain w bits of information*. So far, our considerations are independent of any assumptions about the laws that determine the dynamical trajectories that are physically realized. We can now consider the number of possible trajectories for a given class of *fixed deterministic laws*.

A natural way to define a "class" of laws in CA is by their locality. For example, elementary cellular automata (ECA) are some of the simplest one-dimensional CA studied and are defined by a nearest-neighbor interaction neighborhood for each cell, where cell states are defined on the two-bit alphabet $\{0, 1\}$. Nearest-neighbor interactions define a neighborhood size $L = 3$ (such that a cell is updated by the fixed rule, according to its own state and two nearest neighbors), which we define as the locality of an ECA rule. For ECA there are $R = 2^{2^L} = 256$ possible fixed "laws" (ECA rules) (see Wolfram, 2002). We can therefore set an upper bound for the number of trajectories contained within *any* given rule set, defined by its neighborhood L as:

$$f_L \leq 2^{2^L} \times 2^w \tag{2.3}$$

where f_L is the total number of possible realized trajectories that *any* class of *fixed*, deterministic laws operating with a locality L could realize, starting from a set of 2^w possible initial states (i.e., any initial state with w bits of information).[2] Eqs. 2.2 and 2.3 are not particularly illuminating taken alone; instead we can consider the ratio of the upper bound on the number of trajectories possible

[2] This is an upper bound, as it assumes each trajectory in the set is unique, but as it happens it is possible, for example, that the application of two different ECA rules in the set of all ECA could yield the same trajectory, so this constitutes an absolute upper bound.

under a given set of laws to the total number of deterministic trajectories:

$$\frac{f_L}{N} = \frac{2^{2^L} \times 2^w}{e\Gamma(1+2^w,1)-1} \tag{2.4}$$

Taking the limit as the system size tends toward infinity, that is, as $w \to \infty$, yields:

$$\lim_{w\to\infty}\left[\frac{2^{2^L} \times 2^w}{e\Gamma(1+2^w,1)-1}\right] = 0 \tag{2.5}$$

Note that this result is *independent* of *L*. *For any class of fixed dynamical laws that one chooses (any degree of locality), the fraction of possible physically realized trajectories rapidly approaches zero as the number of bits in states of the world increases*.[3] Thus, the set of all physical realizations of universes evolved according to local, fixed deterministic laws is very impoverished compared with what could potentially be permissible over all possible deterministic trajectories (and worse so if one considers adding stochastic or nondeterministic trajectories in the summation in Eq. 2.2). Only an infinitesimal fraction of paths are even realizable under *any* set of laws, let alone under a particular law drawn from any set.

If we impose time-reversal symmetry on the CA update rules, by analogy with the laws of physics, there is an additional restriction: a small subset of the 256 ECArules are time-reversal invariant. For these laws, there is no single trajectory that includes all possible states (see Figure 2.1). Thus, we encounter the problem that "you can't get there from here," and even if you are in the right regime of configuration space, there is only one path (ordering of states) to follow.

Of course, explanations referring to real biological systems differ from CA in several respects, not least of which is that one deals with macrostates rather than microstates. However, the conclusion is unchanged if one considers the dynamics of macrostates rather than microstates: the trajectories among all possible macrostates

[3] This gets worse if states contain more information, that is, if the alphabet size $m > 2$.

will also be diminished relative to the total number of trajectories (this is because the information in macrostates is less than in the microstates, e.g., will be $< 2^w$ macrostates for our toy example).[4] This toy model cautions us that in seeking to explain a complex world that is ordered in a particular way (e.g., contains living organisms and conscious observers), based on fixed laws that govern microstate evolution, we may well need to fine-tune not only the initial state but also the laws themselves in order to specify the particular ordering of states observed (constraining the universe to a unique past and future). Expressed more succinctly, if one insists on attributing the pathway from mundane chemistry to life as the outcome of fixed dynamical laws, then (our analysis suggests) those laws must be selected with extraordinary care and precision, which is tantamount to intelligent design: it states that "life" is "written into" the laws of physics ab initio. There is no evidence at all that the actual known laws of physics possess this almost miraculous property.

The way to escape from this conundrum – that "you can't get anywhere from here" – is clear: we must abandon the notion of fixed laws when it comes to living and conscious systems.

LIFE ... FROM NEW PHYSICS?

Allowing both the states *and the laws* to evolve in time is one possibility for alleviating the problems associated with fine-tuning of initial states and dynamical laws, as discussed in the previous section. But this cannot be done in an ad hoc way and still be expected to be consistent with our known laws of physics. A trivial solution would be to assume that the laws are time dependent and governed by meta-laws, but then one must explain both where the laws and meta-laws come from, and whether there are meta-meta-laws that govern the meta-laws, and meta-meta-meta-laws, ad infinitum. This is therefore no better an explanation than our current framework,

[4] This holds even if one considers that the number of possible partitions of our state space for 2^W possible states is given by the Bell number B_n, where $n = 2^W$, which approaches ∞ more slowly than the denominator in Eq. (2.4).

thermodynamics, which is a branch of physics[5] due to the mathematical relationship between Shannon and Boltzmann entropies. Substantial work over the last decade has attempted to make this connection explicit; we point the reader to Lutz and Ciliberto (2015) and Parrondo et al. (2015) for recent reviews. Schrödinger was aware of this link in his deliberations on biology, and famously coined the term "negentropy" to describe life's ability to seemingly violate the second law of thermodynamics.[6] Yet he felt that something was missing, and that thermodynamic considerations alone are insufficient to explain life (Schrödinger and Schroedinger, 2004):

> living matter, while not eluding the "laws of physics" as established up to date, is likely to involve "other laws of physics" hitherto unknown …

We suggest one approach to get at these "other laws" is to focus on the connection between the concept of "information" and the equally ill-defined concept of "causation" (Davies and Walker, 2016; Kim et al., 2015; Walker et al., 2016). Both concepts are implicated in the failure of our current physical theories to account for complex states of the world without resorting to very special initial conditions. In particular, we posit that the manner in which biological systems implement state-dependent dynamics is by utilizing information encoded *locally* in the current state of the system, that is, by attributing causal efficacy to information. It is widely recognized that coarse-graining (which would define the relevant "informational" degrees of freedom) plays a foundational role in how biological systems are structured (Flack et al., 2013), by defining the biologically relevant macrovariables (see, e.g., Chapters 10, 12, and 16). However, it is not clear how those macrostates arise, if they are objective or subjective

[5] Stating that information theory is not a physical theory is not the same as saying that information is not physical – a key insight of information theory is that information is a measurable physical quantity. "Information is physical!" in the words of Rolf Landauer (1996).

[6] "Schrödinger's paradox" with regard to life's ability to generate "negative entropy" is quickly resolved if one considers that living systems are open to an environment.

(Shalizi and Moore, 2003), or whether they are in fact a fundamental aspect of biological organization – *intrinsic to the dynamics* (i.e., such that macrostates are causal) rather than merely a useful phenomenological descriptor. A framework in which coarse-grained information-encoding macrostates are causal holds promise for resolving many of the problems discussed herein. This is the key aspect of the hard problem of life.

CONCLUSIONS

There are many difficult open problems in understanding the origin of life – such as the "tar paradox" (Benner, 2014) and the fact that prebiotic chemistry is just plain hard to do. These problems differ qualitatively from the "hard problem of life" as identified here. Most open problems associated with life's origin such as these, while challenging right now, will likely ultimately reduce to known principles of physics and chemistry, and therefore constitute, by our definition, "easy problems." Here we have attempted to identify a core feature of life that won't similarly be solved based on current paradigms – namely, that life seems distinct from other physical systems in how information affects the world (i.e., that macrostates are causal). We have focused on the problem of explaining the pathway from nonliving chemical systems to life and mind to explicate this problem and have attempted to motivate why new principles and potentially even physical laws are necessary. One might regard this as too a radical step; however, it holds potential for resolving deep issues associated with what life is and why it exists. Previous revolutions in our understanding of physical reality, such as general relativity and quantum mechanics, dramatically reshaped our understanding of the world and our place in it. To quote Einstein, "One can best feel in dealing with living things how primitive physics still is" (letter tio L. Szilard, quoted in Prigogine and Stengers, 1997). Given how much more intractable life seems, we should not immediately jump to expecting anything less of a physical theory that might encompass it. If we are so lucky as to stumble on new fundamental understanding of life, it could be such a radical departure from what we know now that it might be left to the next generation of physicists to reconcile the unification of life with other domains of physics, as we are now struggling to accomplish with unifying general relativity and quantum theory a century after those theories were first developed.

ACKNOWLEDGMENTS

This work was made possible through support of a grant from Templeton World Charity Foundation. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of Templeton World Charity Foundation.

REFERENCES

Adams, A., Zenil, H., Davies, P. C. W., and Walker, S. I. (2016). Formal definitions of unbounded evolution and innovation reveal universal mechanisms for open-ended evolution in dynamical systems. *arXiv:1607.01750*.

Benner, S. A. (2014). Paradoxes in the origin of life. *Origins of Life and Evolution of Biospheres*, **44**(4):339.

Campbell, D. T. (1974). Downward causation in hierarchically organised biological systems. In *Studies in the philosophy of biology*, pages 179–186. Springer.

Carter, B., and McCrea, W. H. (1983). The anthropic principle and its implications for biological evolution [and discussion]. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, **310**(1512):347–363.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, **2**(3):200–219.

Cronin, L., and Walker, S. (2016). Beyond prebiotic chemistry. *Science*, **352**:1174–1175.

Darwin, C. (1859). *On the origin of species by means of natural selection*. Murray.

Davies, P. (2008). *The Goldilocks enigma: why is the universe just right for life?* Houghton Mifflin Harcourt.

Davies, P. C. W., and Walker, S. I. (2016). The hidden simplicity of biology: a key issues review. *Rep. Prog. Phys.* **79**(10):102601.

Deutsch, D. (2013). Constructor theory. *Synthese*, **190**(18):4331–4359.

Farnsworth, K. D., Nelson, J., and Gershenson, C. (2013). Living is information processing: from molecules to global systems. *Acta Biotheoretica*, **61**(2):203–222.

Flack, J., Erwin, D., Elliot, T., and Krakauer, D. (2013). Timescales, symmetry, and uncertainty reduction in the origins of hierarchy in biological systems. In Kim Sterelny, Richard Joyce, Brett Calcott, and Ben Fraser (eds), *Cooperation and its evolution*, pages 45–74. MIT Press.

Godfrey-Smith, P., and Sterelny, K. (2008). Biological information. In *The Stanford Encyclopedia of Philosophy* (summer 2016 edition), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/sum2016/entries/information-biological/.

Goldenfeld, N., and Woese, C. (2011). Life is physics: evolution as a collective phenomenon far from equilibrium. *Annu. Rev. Condens. Matter Phys.*: 375–399.

Hofstadter, D. R. (1979). *Godel Escher Bach*. Basic Books.

Jablonka, E., and Szathmáry, E. (1995). The evolution of information storage and heredity. *Trends in Ecology & Evolution*, **10**(5):206–211.

Kim, H., Davies, P., and Walker, S. I. (2015). New scaling relation for information transfer in biological networks. *Journal of the Royal Society Interface*, **12**(113):20150944.

Landauer, R. (1996). The physical nature of information. *Physics Letters A*, **217**(4):188–193.

Lartigue, C., Glass, J. I., Alperovich, N., Pieper, R., Parmar, P. P., Hutchison, C. A., Smith, H., and C., (2007). Genome transplantation in bacteria: changing one species to another. *Science*, **317**(5838):632–638.

Lutz, E., and Ciliberto, S. (2015). Information: From Maxwell's demon to Landauer's eraser. *Physics Today*, **68**(9):30–35.

Monod, J. (1974). *On chance and necessity*. Springer.

Parrondo, J., Horowitz, J., and Sagawa, T. (2015). Thermodynamics of information. *Nature Physics*, **11**:131–139.

Pavlic, T., Adams, A., Davies, P., and Walker, S. (2014). Self-referencing cellular automata: a model of the evolution of information control in biological systems. *arXiv:1405.4070*.

Peirce, C. S. (1982). *Writings of Charles S. Peirce: a chronological edition, volume 1: 1857–1866*, volume 4. Indiana University Press.

Prigogine, I., and Stengers, I. (1997). *The end of certainty*. Simon and Schuster.

Schrödinger, E., and Schroedinger, E. (2004). With mind and matter and autobiographical sketches. In *What is life*, Cambridge University Press, Cambridge UK. Reprinted 2012.

Shalizi, C., and Moore, C. (2003). What is a macrostate? Subjective observations and objective dynamics. *arXiv preprint cond-mat/0303625*.

Smith, E. (2008). Thermodynamics of natural selection. I: Energy flow and the limits on organization. *Journal of Theoretical Biology*, **252**(2):185–197.

Smolin, L. (2013). *Time reborn: from the crisis in physics to the future of the universe*. Houghton Mifflin Harcourt.

Szathmary, E., and Maynard Smith, J. (1994). The major evolutionary transitions. *Nature*, **374**:227–232.

Walker, S., and Davies, P. (2013). The algorithmic origins of life. *J. R. Soc. Interface*, **10**(79):20120869.

Walker, S. I. (2015). Is life fundamental? In Aguirre, A., Foster, B., and Merali, Z., editors, *Questioning the foundations of physics: which of our fundamental assumptions are wrong?* Springer: 259–268.

Walker, S. I. (2016). The descent of math. In Aguirre, A., Foster, B., and Merali, Z., editors, *Trick of truth: the mysterious connection between physics and mathematics?* Springer: 183–192.

Walker, S. I., Cisneros, L., and Davies, P. C. W. (2012). Evolutionary transitions and top-down causation. *arXiv preprint arXiv:1207.4808*.

Walker, S. I., Kim, H., and Davies, P. C. W. (2016). The informational architecture of the cell. *Phil. Trans. A*, page 20150057.

Webb, J. K., King, J. A., Murphy, M. T., Flambaum, V. V., Carswell, R. F., and Bainbridge, M. B. (2011). Indications of a spatial variation of the fine structure constant. *Physical Review Letters*, **107**(19):191101.

Wheeler, J. A. (1983). On recognizing without law, oersted medal response at the Joint APS-AAPT meeting, New York, 25 January 1983. *American Journal of Physics*, **51**(5): 398–404.

Whitesides, G. (2012). The improbability of life. In Barrow, J., Morris, S., Freeland, S., and Harper, C. Jr., editors, *Fitness of the cosmos for life*, volume 1: xi–xii. Cambridge: Cambridge University Press.

Wolfram, S. (2002). *A new kind of science*, volume 5. Wolfram Media.

a constructor, and the chemicals being transformed are its substrates. By *attribute* here one means a set of states of a system in which the system has a certain property according to the subsidiary theory describing it – such as being red or blue. The basic entities of constructor theory are *tasks*, which consist of the specifications of only the input–output pairs of a transformation, with the constructor abstracted away:

Input attributes of substrates → Output attributes of substrates.

Therefore, a task A on a substrate **S** is a set:

$$\mathrm{A} = \{x_1 \to y_1,\ \ x_2 \to y_2, \ldots\},$$

where the $x_1, x_2, \ldots$ and the $y_1, y_2, \ldots$ are attributes of **S**. The set $\{x_i\} = In(\mathrm{A})$ are the legitimate input attributes of A and the set $\{y_i\} = \mathrm{Out}(\mathrm{A})$ its legitimate output attributes. Tasks may be composed into networks, by serial and parallel composition, to form other tasks.

Quite remarkably, this is an explicitly local framework, requiring that individual physical systems have states (and attributes) in the sense described. Indeed, another cardinal principle of constructor theory is Einstein's principle of locality (Einstein, 1949): *There exists a mode of description such that the state of the combined system* $\mathbf{S}_1 \oplus \mathbf{S}_2$ *of any two substrates* $\mathbf{S}_1$ *and* $\mathbf{S}_2$ *is the pair* (x, y) *of the states* x *of* $\mathbf{S}_1$ *and* y *of* $\mathbf{S}_2$*, and any construction undergone by* $\mathbf{S}_1$ *and not* $\mathbf{S}_2$ *can change only* x *and not* y*.* In quantum theory, the Heisenberg picture is such a mode of description (see Deutsch, 2000).

A constructor is *capable of performing a task* A if, whenever presented with substrates having an attribute in *In*(A), it delivers them with one of the corresponding attributes in *Out*(A). A task A is *impossible* if it is forbidden by the laws of physics. Otherwise it is *possible* – which means that the laws of nature impose no limit, short of perfection, on how accurately A could be performed, nor on how well things that are capable of approximately performing it could retain their ability to do so again. However, it is crucial to bear in mind that *no perfect constructors exist in nature*, given our laws of physics. Approximations to them, such as catalysts, living cells, or robots, do make errors and also deteriorate with use. However, when a task is possible, the laws of nature permit the existence of an approximation to a constructor for that task to any given finite accuracy. The notion of a constructor is shorthand for the infinite sequence of these approximations.

Therefore, in this framework a task either is categorically impossible or is possible. Both are *deterministic* statements: in the worldview of constructor theory, probabilistic theories can only be approximate descriptions of reality. Probabilities are indeed emergent in constructor theory – see Marletto (2015b) – just like in unitary quantum theory (Deutsch, 2000; Wallace, 2003). For how such a theory can be testable, see Deutsch (2015).

Although a task refers to an isolated system of constructor and substrates, one is sometimes interested in what is possible or impossible irrespective of the kind of resources required. So if it is possible to perform the task A in parallel with some task T on some generic substrate that is preparable – see Deutsch and Marletto (2015) – one says that A is *possible with side effects*, which we write as $\mathrm{A}^{\underline{\checkmark}}$. The task T represents the side effect.

So in constructor theory everything important about the world is expressed via statements about the possibility and impossibility of tasks, *without mentioning constructors*. One might wonder what difference switching to this formulation can possibly make. After all, it is perfectly possible to express the possibility or impossibility of a task in the prevailing conception, as a conditional statement about the composite system of the constructor and the substrates, given certain initial conditions and the laws of motion. However, as we are about to see, the constructor-theoretic approach (where one can abstract away the constructor) makes all the difference in the case of information.

Whether or not information is physical (Landauer, 1961) and what this can possibly mean has been at the centre of a long-standing debate. Information is widely used in physics, but appears to be very different from all the entities appearing in the physical descriptions of the world. It is not, for instance, an observable – such as the position or the velocity of a particle. Indeed, it has properties like no other variable or observable in fundamental physics: it behaves like an abstraction. For there are laws about information that refer directly to it, without ever mentioning the details of the physical substrates that instantiate it (this is the *substrate-independence* of information), and moreover it is *interoperable* – it can be copied from one medium to another without having its properties qua information changed. Yet information can exist only when physically instantiated; also, for example, the information-processing abilities of a computer depend on the underlying physical laws of motion, as we know from the quantum theory of computation. So, there are reasons to expect that the laws governing information, like those governing computation, are laws of physics. How can these apparently contradictory aspects of information be reconciled?

The key to the answer is that the informally conceived notion of information implicitly refers to *certain interactions being possible* in nature; it refers to the existence of certain regularities in the laws of physics. A fundamental physical theory of information is one that expresses such regularities. As an example of what these regularities are, consider interoperability – as we said above, this is the property of information being *copiable* from one physical instantiation (e.g., transistors in a computer) to a different one (e.g., DNA). This is a regularity displayed by the laws of physics of our universe, which is taken for granted by the current theories of information and computation. However, one could imagine laws that did not have it – under which 'information' (as we informally conceive it) would not exist. For example, consider a universe where there exist two sectors A and B, each one allowing copying-like interactions between media inside it, but such that no copying interactions were allowed between A and B. There would be no 'information' (as informally conceived) in the composite system of the two sectors. This is an example of how whether or not information can exist depends on the existence of certain regularities in the laws of physics. These regularities have remained unexpressed in fundamental physics; the constructor theory of information precisely expresses them in the form of new, conjectured laws of physics. This is how information can be brought into fundamental physics: one does so by expressing in an exact, scale-independent way what constraints the laws of physics must obey in order for them to instantiate what we have learnt informally to call 'information'.

It is not surprising that constructor theory proves to be particularly effective to this end. As Shannon and Weaver (1949) put it, information has a counterfactual nature:

> this word 'information' in communication theory relates not so much to what you do say, as to what you could say.

The constructor theory of information differs from previous approaches to incorporating information into fundamental physics, e.g. Wheeler's 'it from bit' (Wheeler, 1990), in that it does not consider information itself as an a priori mathematical or logical concept. Instead, it requires that the nature and properties of information follow entirely from the laws of physics.

The logic of how the theory is constructed is elegant and simple. The first key step is that in constructor theory information is understood in terms of computations, not vice versa as is usually done. So, first one defines a *reversible computation* $\mathrm{C}_{\Pi}\,(S)$ as a task – that of performing, with or without side effects, a permutation Π over some set S of at least two possible attributes of some substrate:

$$\mathrm{C}_{\Pi}\,(S) = \bigcup_{x \in S} \{x \to \Pi(x)\}\ .$$

By a 'reversible computation' C_{Π} is meant a logically reversible, i.e., one-to-one, task, but the process that implements it may be physically irreversible, because of the possibility of side effects. A *computation variable* is a set S of two or more possible attributes for which $\mathrm{C}_{\Pi}^{\underline{\checkmark}}$ for all permutations Π over S, and a *computation medium* is a substrate with at least one computation variable. A quantum bit in any two

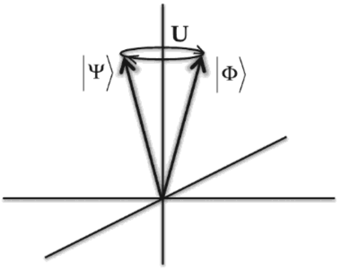


FIG. 3.1 Pictorial representation of the swap U of two nonorthogonal states $|\Psi\rangle$ and $|\Phi\rangle$ of a quantum bit (e.g., a $\frac{1}{2}$ spin). U is unitary as prescribed by quantum theory, i.e., it preserves the inner product; applying it twice is equivalent to the identity.

nonorthogonal states is an example of a computation medium: these two states can be swapped by a unitary transformation, despite their not being distinguishable by a single-shot measurement (Deutsch and Marletto, 2014). See Figure 3.1.

The next step is to introduce the notion of an *information medium*, which requires one to consider computations involving two instances of the same substrate **S**. The *cloning task* for a set S of possible attributes of **S** is

$$\mathrm{R}_S\,(x_0) = \bigcup_{x \in S} \{(x, x_0) \to\ (x, x)\} \tag{3.1}$$

on **S** $\oplus$ **S**, where x_0 is some fixed (independent of x) attribute with which it is possible to prepare **S**. A set S is *clonable* if R_S is possible (with or without side effects) for some such x_0.

An *information variable* is then defined as a clonable computation variable. An *information attribute* is one that is a member of an information variable, and an *information medium* is a substrate that has at least one information variable.

We are now in the position to express exactly what it means that a system contains information. In particular, a substrate **S** *instantiates information* if it is in a state belonging to some attribute in some information variable S of **S** and if it could have been given any of the other attributes in S. The constructor-theoretic mode of explanation communication scheme investigated by Shannon (as discussed in Deutsch and Marletto, 2015). This theory also provides a framework, independent of quantum theory, where one can investigate information under a broad range of theories (including 'postquantum' theories) if they obey the principles. Finally, it provides an exact, scale-independent physical characterisation of what it means for laws of physics to allow for 'information'.

CONSTRUCTOR THEORY OF LIFE

The constructor theory of life (Marletto, 2015a) builds on this to tackle a problem relevant to the foundations of both physics and biology. To understand what the problem is, let me first clarify the connection between evolutionary theory and fundamental physics.

The problem the theory of evolution was proposed to address is the existence of living entities. That living entities are particularly remarkable physical objects has been known since antiquity. They struck the curiosity of early humans, as the superb example of cave paintings testifies (see Wagner and Briggs, 2016). Early attempts to classify the properties of living things, to distinguish them from inert matter, date back to Socrates and Aristotle. However, only in modern times was it possible to express the objective property that makes their existence a *scientific problem*. This property is rooted in physics.

In modern biology, living entities are characterised by the *appearance of design* displayed in their biological adaptations. As William Paley put it, they have *several, different subparts all coordinating to the same purpose* (Paley, 2006). For instance, trunks in elephants appear as highly designed objects serving a specific function.

Darwin's *theory of evolution* was proposed precisely to explain how the appearance of design in living things can have come about without an intentional design process, via the undesigned process of variation and natural selection. It is notable that Darwin's theory is based on an (informal) constructor-theoretic reasoning. Despite its predictive power – see, e.g., the famous case of Darwin's moth (Kritsky, 1991) – the core statement of the theory is not in the form of a prediction. It does not state that, say, elephants' trunks must exist; it states that it is *possible* that living things have come about via natural selection and variation, without an intentional design process, and explains how. Recasting this statement in terms of predictions would not serve the purpose. Having a prediction (probabilistic or not) that maintains that, say, elephants will occur (or will probably occur) at some point in our universe does not rule out the possibility that the laws or the initial conditions contain design and thus would not serve the purpose of understanding how elephants can have come about without an intentional design process.

In constructor theory, one can express more precisely how the problematicity of living things is rooted in physics: living things are problematic because, in sharp contrast with inert matter, they *approximate accurately the behaviour of a constructor*. They perform tasks to a *high accuracy*, *reliably*, and they maintain this property in time, displaying *resiliency*. This is problematic because of how we conjecture the laws of physics of our universe are: under such laws, the *generic resources* – defined as the physical objects that occur in effectively unlimited numbers (such as atoms, photons, and simple chemicals) – are elementary. In particular, they do not contain accurate constructors: if they ever perform tasks, they do so only to a low, finite accuracy. Moreover, under such laws it is impossible that an accurate constructor arises from generic resources only, acted on by generic resources only. I shall call laws of this kind *no-design laws* (Marletto, 2015a).

Thus the problem about the existence of living entities – the problem that the theory of evolution aims at solving – is better expressed as that of how accurate constructors such as living entities *can* emerge from generic resources, without an intentional design process, given laws of physics that are no-design. This reveals the connection between evolutionary biology and fundamental physics.

In the modern *neo-Darwinian synthesis* (Dawkins, 1976, 1999), the explanation of Darwin's theory was merged with molecular biology, where the connection with physics becomes more explicit. The centerpiece of the explanation is a physical object, the *replicator*, that is copied in the following pattern:

$$(R, N) \overset{C}{\rightarrow} (R, R, W)$$

where R is the replicator and C is a constructor for the copying (a copier), acting on some generic resources N (possibly producing waste products W).

In nature this process occurs to different accuracies. Short RNA strands and simple molecules involved in the origin of life (Szathmáry and Maynard Smith, 1997) are poor, inaccurate replicators. In those cases, the copier C is implicit in the dynamical laws of physics. In cells, at the other extreme, the replicator R is the DNA strand, which is copied very accurately by various enzymes.

The replication of replicators in cells relies crucially on the ability of a cell to undergo *accurate self-reproduction* – the construction where an object S (the *self-reproducer*) brings about another instance of itself, in the schematic pattern:

$$(S, N) \rightarrow (S, S, W)$$

Here W represents products; the raw materials N do not contain the means to create another S; and the whole system could be isolated. Thus a self-reproducer S cannot rely on any mechanism other than itself to cause the construction of another S, unlike a replicator R that is allowed to use an external copying mechanism, such as C.

That evolutionary theory relies on both these processes being possible constitutes the problem addressed by the constructor theory of life. Indeed both replication and self-reproduction, which is essential to replication, occur in living things with remarkable accuracy. Thus it is necessary, for the theory of evolution fully to explain the appearance of design in the presence of no-design laws, to provide an additional argument of how and why these processes are compatible with underlying dynamical laws that are no-design, i.e., that do not contain the design of biological adaptations. The constructor theory of life provides precisely this explanation.

The compatibility of accurate self-reproduction with the laws of physics has indeed been contested, along the lines advocated by Wigner, who proposed the claim that accurate self-reproduction, as it occurs in living cells, requires laws of physics that are 'tailored for self-reproduction to occur' (Bohm, 1969; Wigner, 1961).

Wigner uses technical quantum theory to make his case. But his claim is actually simpler and broader than that and expresses the problem: how can self-reproduction be so accurate in the presence of no-design laws – laws of physics that are simple and do not contain any reference to accurate self-reproducers or replicators? His statement would have wide implications were it true. Not only would it require our laws of physics to be complemented with ad hoc ones, containing the design of biological adaptations, but also the theory of evolution would, after all, rely on laws of physics containing the design of biological adaptations.

The constructor theory of life shows that accurate self-reproduction and accurate replication are possible under no-design laws, thus rebutting that claim and vindicating the compatibility of Darwin's theory of evolution with no-design laws. It also shows what other features no-design laws must have to allow those processes; in particular, they must allow the existence of information media, as defined in constructor theory. In addition, it shows what logic accurate self-reproducers must follow, under such laws; it turns out that an accurate replicator must rely on a self-reproducer, and vice versa.

The logic of the argument is as follows. First, one notes that replicators are already expressed naturally in the constructor theory of information; see Equation (3.1): substrates allowing a set of attributes that can be permuted in all possible ways and replicated (i.e., cloned) are information media. Moreover, self-reproducers are characterised as constructors for another instance of themselves, as we can see by rewriting the self-reproduction of a self-reproducer S, in the presence of generic resources N only (with possible waste products W) in a constructor-theoretic notation:

$$N \xrightarrow{S} (S, W)$$

Thus, the problem can be reformulated naturally and exactly, in constructor theory, as: Are accurate self-reproducers and replicators possible under no-design laws?

Furthermore, the appearance of design and the notion of no-design laws can both be expressed, exactly, within constructor theory. Here it is crucial that constructor theory allows one to avoid using the notion of probability to model those concepts. In particular, we are interested here in defining no-design intending the design being *that of biological adaptations*. Laws of physics might be fine-tuned in other senses (see Davies, 2000), but here we are interested only in design of living things. In the prevailing conception, resorting to probabilities, it is not possible to model this concept. For example, one could say that some dynamical law that is nontypical under some natural measure is a designed one, as Wigner conjectured. But clearly this is a non sequitur: the unitary of our universe is indeed 'nontypical' because, e.g., it is local. But this gives no indication as to whether it contains the design of biological adaptations. In constructor theory, instead, one can characterise precisely, within physics and without resorting to probability, what 'no-design laws of physics' are. They are, as expressed above, laws whose generic resources do not contain accurate constructors, nor do they allow the sudden arising of accurate constructors out of generic resources only.

A similar approach allows one to express the appearance of design. The latter also cannot be modelled by being 'improbable'. For probabilities are multiplicative, but the appearance of design of the composite system of two objects with the appearance of design need not have more of an appearance of design than either of the two separately. In constructor theory one can give, instead, an elegant constructor-theoretic expression that has the required property, in terms of *programmable constructors* – constructors that have, among their input substrates, an information medium holding one of its

Since each step is elementary, this process is compatible with no-design laws of physics. Thus, von Neumann's original discovery about the logic of self-reproduction in the purely computational context of cellular automata is extended here to the actual laws of physics. It is shown, in particular, that this logic (formerly proven to be sufficient in that context) is *necessary* for accurate self-reproducers to be possible, under no-design laws of physics, e.g., the ones conjectured to rule our universe. As a result, this also implies that an *accurate replication*, as it occurs in living entities, *requires* a vehicle that can perform error-correction – and thus a self-reproducer. This is an interesting spin-off, subverting the assumption that most neo-Darwinian theorists would take, that 'The only thing that is really fundamental to Darwinian life is self-replicating, coded information – genes, in the terminology of life on this planet' (Dawkins, 1976).

3. The last step of the argument is to explain how it is *possible* that accurate self-reproducers (and accurate replicators) have arisen from naturally occurring resources, under no-design laws. The theory of evolution by natural selection and variation provides the explanation for how this occurs: constructor theory shows that this explanation is indeed compatible with no-design laws, establishing two points. The first one is that the logic of evolution by natural selection and variation operates by nonspecific, elementary steps that are *not systematically directed to improvement*. Indeed, the variations caused by the environment in the populations of replicators on which the selection operates are *nonspecific* to the end product, and natural selection is *blind and undirected*. Indeed, the whole process is a highly inaccurate construction for the emergence of accurate self-reproducers from inaccurate ones, given enough time and resources. This construction is so inaccurate and unreliable that it requires no further explanation, as it is compatible with no-design laws. The second point is that natural selection, *to get started*, does not require accurate self-reproducers to be in the initial generic resources. It is sufficient that the latter contain *only inaccurate ones*, such as short RNA strands capable of achieving highly approximate

replication without a vehicle. This concludes the proof that accurate self-reproducers and replicators are possible under no-design laws.

Note that here the problem was not that of predicting with what probability accurate self-reproducers would arise, given certain initial conditions and laws of motion – a problem that has been tackled in Walker and Davies (2013). The problem was a constructor-theoretic one: whether, and how, accurate self-reproducers are *possible, under no-design laws*, and how accurate they are. This problem can be addressed without explicitly formulating any predictions. The final conclusion is that those accurate constructors are permitted under such laws, provided that these laws allow the possibility of modular, discrete replicators to be physically instantiated. In constructor-theoretic terms, *it is necessary that the laws allow information media*. This is also what Darwin's theory of evolution requires of the laws of physics. Rather crucially, this is a requirement that is nonspecific to life. Note also that this is not the usual claim that a vague notion of information is needed for life. The statement of the constructor theory of life is an exact, scale-independent one, based on the rigorous notion of information media provided by the constructor-theory of information.

The recipe in the self-reproducer can be characterised in constructor theory as a *special kind of information*, which *causes itself to remain instantiated in physical systems and can act as a constructor*. We call it *knowledge*, in the sense of Popper's objective knowledge (Popper, 1992). Notably, not all information acts as a constructor, and not all information that can act as a constructor is knowledge. To explicate the distinction, one can consider the difference between a generic sequence of letters (which simply instantiates information, as it is copiable); a syntactically meaningful, but faulty, computer program (which can act as a constructor when executed, but lacks the ability to cause itself to remain instantiated in physical systems, as it is 'fruitless'); and a computer program implementing an effective algorithm (which does, indeed, instantiate knowledge). This distinction