# Computation, Information, Cognition

# Computation, Information, Cognition
# The Nexus and the Liminal

Edited by

Susan Alice Jane Stuart
& Gordana Dodig Crnkovic

CAMBRIDGE SCHOLARS PUBLISHING

Computation, Information, Cognition: The Nexus and the Liminal, edited by Susan Alice Jane Stuart and Gordana Dodig Crnkovic

This book first published 2007 by

Cambridge Scholars Publishing

15 Angerton Gardens, Newcastle, NE5 2JA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library


ISBN 1-84718-090-6

# TABLE OF CONTENTS

# PREFACE

In this book we have tried to draw together a number of important strands in contemporary approaches to the philosophical and scientific questions that emerge when dealing with the issues of computing, information, cognition and their overlap. It is a work that has come about as a result of establishing a European forum for, what were initially, a series of North American Computing and Philosophy (NA-CAP) conferences whose broad concern was with all aspects of the computational turn that was occurring within the discipline of philosophy. Initially the emphasis was on computer-assisted instruction, but quickly it grew to encompass the possibility of creating minimally self-conscious artificial agents, the ethics of artificial intelligence and machine consciousness, intellectual property rights, issues surrounding perceptual software and questions of representation *versus* non-representation, adaptation and autonomy in robotics, and philosophical issues in the domains of analogue and digital librarianship and archiving. These meetings also saw the development and flourishing of a new perspective, the Philosophy of Information, concerned with the conceptual issues that arise at the intersection of computer science, information technology and philosophy. As Luciano Floridi explains, the Philosophy of Information is concerned with

– the critical investigation of the conceptual nature and basic principles of information, including its dynamics, utilisation and sciences
– the elaboration and application of information-theoretic and computational methodologies to philosophical problems.

["What is the Philosophy of Information", *Metaphilosophy*, 2002, (33), 1/2.]

The first conference of the European Computing and Philosophy (E-CAP) took place in 2003 at the University of Glasgow, Scotland. This was followed in 2004 with a conference at the University of Pavia, Italy, and in 2005 at the Mälardalen University,Västerås, Sweden. It is from the 2005 conference that the papers in this volume are derived.

The presentations in the European, North American, and now Asian-Pacific (AP-CAP) conferences continue to be shaped by the computational turn but, as the fields of enquiry advance and expand, new elements appear and, once again, new perspectives emerge. It is with considerable pleasure that we present this volume not only with a rich opening section dealing with the Philosophy of Information, but also with a section on the Philosophy of Ontology and the ontological considerations involved in the creation of controlled vocabularies and the semantic relations between their terms in software environments. From their work on the development of a programming language ontology (this volume) Eden & Turner have gone on to establish, with an internationally-based group of colleagues, the Philosophy of Computer Science as a fresh and potentially rich perspective from which to address all aspects, physical and metaphysical, of computer science and its objects. The Philosophy of Computer Science was also one of the main themes at E-CAP 2006 at the Norwegian University for Science and Technology, Trondheim, Norway.

So, the background to this volume is exciting being informed by, at least, three new philosophical perspectives, and it is dynamic for the conferences and dialogue have now spread to become truly international.

With so much activity there are a great many people to thank for their vision, influence, good-judgement, and kindness. The first of these must be Robert Cavalier, for without his enthusiasm to see CAP conferences spread to Europe, E-CAP would not have become what it is today. Then there's a clutch of really great people who are a delight to know and work with in the International Association for Philosophy and Computing (IA-CAP), and who can always be relied upon for sound advice, they include but are not limited to, Ron Barnette, Charles Ess, Marvin Croy, and Luciano Floridi.

And, finally, it is essential that we thank the authors for the excellent collaboration and everyone who reviewed submissions for the E-CAP05 conference and to those who helped with the subsequent tough selection process for papers most appropriate for the book. In the latter category we must include Peter Århem, Gustaf Arrhenius, Birgitta Bergsten, Rikard Bonner, Søren Brier, Göran Collste, Chris Dobbyn, Kaj Börje Hansen, Lars-Göran Johansson, Torbjörn Lager, Staffan Larsson, Pedro C. Marijuán, Christina Mörtberg, Joakim Nivre, Jan Österberg, Bertil Rolf, May Thorseth, and Tom Ziemke.

All that is left to say is that we hope you enjoy reading this book and are as stimulated by the papers as we have been.

Susan A. J. Stuart and Gordana Dodig-Crnkovic

March 2007

# INTRODUCTION

SUSAN STUART
AND GORDANA DODIG-CRNKOVIC

Every epoch and culture has a different conception of the Universe. For Ptolemy, Descartes, and Newton the Universe was best conceived in a mechanistic way as some vast machine. For others it is, in its entirety, a living organism. [*Viz*. Thales of Miletus, Spinoza, and Kafatos & Nadeau 1999.] Our current understanding in terms of information and computing has led to a conception of the Universe as, more or less explicitly, a computer. On such a pancomputational and paninformational view (Zuse 1967; Lloyd 2006; Chaitin this volume ), if all physics is expressible as computation – so the whole universe can be represented as a network of computing processes at different levels of granularity – then we can consider information as a result of (natural) computation, and the Universe as a network of computing processes that are defined by the information they manipulate and produce. Under this conception information is that which constitutes the computing structure, the Universe, at any given moment; the structure changes continuously and that change can be understood as computation. Thus, computation is what happens dynamically to information from one moment to the next.

Under this conception information and computation are elements of a dual-aspect theory, providing a dichotomy with which we can begin to grasp the more fundamental classical energy / matter dichotomy that, embedded within space / time, is generally taken to constitute reality. The elements in such pairings are not only complementary, they are also interdependent, like discreteness and continuum, like time and space, like wave and particle, like form and content; computation and information presuppose one another and can only be understood fully in their conjunction. (Dodig-Crnkovic 2006) On the semantic level, that there might be a wealth of computational processes underlying our communication is insufficient if there is nothing to communicate; that there might be information to be communicated is insufficient without the computational process. One is reminded very clearly of

Kant's dictum: "Thought without content is empty, intuitions without concepts are blind" [A52/B76].

In expanding its domain to the whole Universe computation goes well beyond the limits set by the Church-Turing computability hypothesis. Being process-oriented, pancomputationalism is not so much interested in the structure of the Universe, that is, whether it is ultimately digital, analogue, some hybrid of the two, or none of these, for it takes structure as a given and develops its dynamical behaviour over time. In this way, pancomputationalism in its most general form, natural computation, works equally well for both digital and analogue computing processes.

But, if we are interested in structures and adopt an information realism where entities are informational (including semiotic) structures – the first view we can see expressed in the biological work of Marijuán & Moral and the second in Brier in this volume – then we can opt instead for information structuralism. (Floridi this volume). With regard to this position, Floridi (2004) identifies information semantics as one of the Open Problems in the Philosophy of Information. Among the advantages the informational view brings with it is its finer granularity when put into context of wisdom, knowledge, information, and data. In the adoption of this view we must concern ourselves with the dynamics of information and, more specifically, with whether there is an *information logic* (IL) that is distinct from *epistemic logic* (EL) and *doxastic logic* (DL). It is with this particular issue that Floridi is concerned in his article in this volume; his fundamental claim is that being informed is a transitive state, but that knowing and believing are not.

The pursuit and acquisition of knowledge, and thus also information, has frequently been thought to be a double-edged sword. We seek knowledge so that we can progress in our culture, our science and our technology, but there are times when our discoveries can have a Pyrrhic feel to them In the Abrahamic traditions gaining knowledge is associated with expulsion from paradise, with the loss of safety and certainty. However, in being expelled from one Paradise, we search for or create new ones; in our abandonment of the Ptolemaic astronomy of ideal spheres embrace the Newtonian-Laplacean perfection of heavenly clockwork.

In an echo of the Genesis story and its sentiment Hilbert (1926) said "No one will drive us from the paradise which Cantor created for us". But, one by one our once firmly held convictions are having to be abandoned. We have been removed from our position as the absolute centre of the Universe with its unique and privileged system of co-ordinates, and now find ourselves in the outskirts of our galaxy, a galaxy which is in no way special amongst galaxies. We have left behind absolute space and absolute time. It is time to leave the absolute truth of there being one, and only one, true system of logic (logical monism). Logical pluralism (Beall & Restall, 2000) is motivated by an analysis of disagreement within the classical first-order logic, relevant logic and intuitionistic logic in the account of logical consequence (and hence of logical truth). Allo (this volume) argues that logical pluralism could also entail semantic informational pluralism as information content depends upon the underlying logic one assumes. One of the consequences of this view is that, when a formal account of semantic information is elaborated, the absolute validity of logic cannot be relied upon and some further domain-specific motivations will be required if we are to be assured of the appropriateness of the logic we choose to use.

It is to the external world that we now look for motivations and specifically to the interdisciplinary work being done in the overlapping fields of information, biology and biosemantics, cognitive science, computational linguistics, technology, ontology, ethics, and ultimately – closing the circle – education. We are attempting to draw together some of the most significant work that exists at, what are often conceived to be, the borders of traditional disciplines. In other words we are attempting to establish a nexus in the liminal.

We begin with a speculative metaphysics provided by Greg Chaitin's article 'Epistemology as Information Theory: From Leibniz to the Omega Number'. Chaitin's mix of digital philosophy and digital physics presents a "neo-Pythagorean vision of the world" in which "everything is made out of 0/1 bits, everything is digital software, and God is a computer programmer, not a mathematician!". The challenge is to find how much will fit this new theory, how much will fail to fit, and what it will help us to understand. Chaitin makes the eminently computationalist claim that "we only understand something if we can program it" and, since what we can program depends on the laws of physics that hold in this universe, the things that don't fit the theory will be those things we can't program and, thus, cannot understand. Then he confronts the possibility of uncomputable real numbers by arguing that real numbers don't exist. In support of his position Chaitin presents an algorithmic information theory based on Leibniz's dictum (1686) that the universe has been created simplest in hypotheses and richest in phenomena, thus any explanation has to be simpler than that which it attempts to explain. In terms of digital philosophy and the pancomputational universe this becomes the claim that an elegant program has the property that no program written in the same programming language, that produces the same output, is smaller than it, that is to say: "an elegant program is the most concise, the simplest, the best theory for its output." But what happens if it is not possible to deduce the truth of something from any principle simpler than itself; well, Chaitin concludes, "proofs become useless, because **anything** can be proven from principles that are equally complicated", and this is what happens with any static formal and axiomatic theory for all mathematics. [*Viz.* Hilbert's programme 1920's] There is no one single, simple and basic proof or axiom upon which all else depends; mathematics is more like biology, a world of infinite complexity which cannot be explained in a finite bit theory. However, Chaitin's ambitious programme is not simply to suggest a pluralism of axioms, but rather that, in our attempt to understand mathematics, biology, the universe, we must be willing to add new axioms that need not be self-evident but which can be justified pragmatically. He is a thinker with great energy and panache, and the courage of his ideas is conveyed to the reader very well through his text.

Pancomputationalism, paninformationalism, complexity, axiomatic and logical pluralism, ontology, and Leibniz are some of the leitmotivs that emerge from Chaitin's article and which arise in many of the other contributions in this volume. Pietarinen's article connects with the dichotomy of the continuous and the discrete in the context of an attempt to bridge the current division between analogue and digital accounts of the mind. His emphasis is on visual and non-symbolic representational systems in logic, computing and cognitive sciences, and he examines Peirce's iconic logic – a logic Peirce, himself, thought to be superior to his symbolic system – that employs the diagrammatic logic of Existential Graphs. Graphs may be discrete objects in one sense, but they are also a summation of non-verbalised reasoning processes, in effect, a different, direct form of visual reasoning. Again we are confronted with claims for complexity and a need for an expanded logical system, though this time it is one that is necessary if we are to eventually comprehend the logic of icons and the nature of the non-discrete, analogue mind.

Of perennial concern to philosophy, physics, and now the cognitive sciences, are the notions of causation and causal relations; their importance is no less considerable in the domains of information and computation. "Intuitively there is conceptual connection between causation and transfer of information, because we can't get any information from a system without interacting causally with it ... Thus transfer of information is a causal process.". [Johansson, this volume] Johansson addresses this concern by first establishing four, quite different, proposals: causation is (i) the transfer of a conserved quantity; (ii) analysed in terms of counterfactuals; (iii) explicable in terms of INUS-conditions, that is, a cause can be an insufficient [I] but necessary [N] part of a condition which is itself unnecessary [U] but exclusively sufficient [S]; and (iv) something we humans – agents – can manipulate in, for example, the transfer of information which can be a cause of something else. Johansson's suggestion is that, in an attempt to understand this rather complicated notion of causality, we must drop the counterfactual approach – it leads us up a blind alley when we try to determine its truth-value – and unify the other three approaches to provide a robust concept of cause as it is used in ordinary language, and in the natural and social sciences.

In Gang Liu's essay we return to Leibniz as a key figure in the convergence of computing and philosophy, but this time as a proto-Sinologist who interprets the *I-Ching* hexagrams and attempts to integrate elements of oriental organic philosophy into his metaphysics and especially into the semantics of possible worlds. Taking this as his starting point Liu presents a new synthesis with modal information theory (MIT) or modal informationalism in an effort to deal with the indecisive ontological status of information and its shift within the idealism / materialism dichotomy. He closes his essay by urging us to re-examine Leibniz's work in the context of the philosophy of information and pancomputationalism: "A re-discovery of Leibniz's philosophy is essential; his ideas might have been too radical to be accepted in his own age, let them not be too radical to be accepted by ours, especially with the advent of a new field, the philosophy of information."

From questions about the ontological status of information *per se* we move to the ontology of controlled vocabularies with Smith & Ceusters; we find, even here, Leibniz's influence, though this time it is his vision of a universal language, a *mathesis universalis.*

Now that all states, transactions, and relations – in business, medicine, administration, astronomy, and so on – are to be formalised to enable automatic communication and sharing, we must solve the problems of ambiguity and indeterminacy common to all concepts embedded in natural language. The creation of appropriate ontologies confronts this problem head on. "Ontology, as conceived from the realist perspective, is not a software implementation or a controlled vocabulary. Rather, it is a theory of reality, a '*science of what is, of the kinds and structures of objects, properties, events, processes and relations in every area of reality*' (Smith 2003)." In their essay 'Ontology as the Core Discipline of Biomedical Informatics. Legacies of the Past and Recommendations for the Future Direction of Research' Smith & Ceusters provide an eloquent account of the misapplication of controlled vocabularies which provide an inflexible uniform framework within which it is supposed that items and their relations can be adequately described. The prominent example of a controlled vocabulary is the Gene Ontology which, in its use of the very limited 'is-a' and 'part-of' relations, runs counter to the rules of logic. They identify one very telling failure, when 'is-a' is used tacitly to mean 'within'; for example, when wanting to express the fact that the embryo is in the uterus the gene ontologist writes completely falsely that the embryo 'is-a' uterus. Their paper, whilst wryly amusing, is a very stark warning about the urgent need for a coherent ontology, and not just for biomedical informatics.

In a very similar vein Munn addresses the ontologies, the taxonomical hierarchies, of biomedical information systems, but her focus is on the use of the term 'function'. In her analysis of the National Library of Medicine's clinical resource, Medical Subject Headings (MeSH), she discovers that 'function' is used to mean 'role' and that 'role' has two characteristics: it is something carried out in relation to a biological system, and it is associated with repeated patterns of action. These characteristics presuppose that function is normative and, thus, something which provides a standard from which we can judge *with respect to what* conditions *X* functions well. It is these normative standards that bioinformatics must specify if it is to be successful.

Chaitin's computationalist dictum, that you can only really understand something if you can program it, is clearly evident in each of the ontology papers in this volume.

In Eden & Turner's essay we have an analysis of the ontology of programming languages and subsequent possibility of a philosophical analysis of software design. They begin by considering the different possible ontological perspectives and settle on a *denotational semantics* (DS) as "the most ontologically self-conscious", it being the one that makes explicit what set theoretic constructions are required to interpret the language in set theory. Doing this, they argue, will lead back to a Quinean distinction between the ontological commitments of a scientific theory and the actual choice of a theory. They examine the assumptions of compositionality, extensionality and completeness that are the underpinning of DS, and question the set-theoretic account of programming languages which accepts sets as a basic ontological structure. Like constructive logicians and mathematicians who appeal to notions such as *operation* and *constructive proof*, computer scientists, they argue, have their own basic ontology in *data types*.

Again in the context of the relation of information to knowledge Hagengruber & Riss address the issue of knowledge management, the problem of transforming human knowledge into machine representations, and knowledge domain extensionality and integration. In particular, their focus is on the dynamic nature of the switch from implicit to explicit knowledge and, since their view of knowledge is related to action, they adopt a dynamic approach. Critically, they ask why is it possible for us to put together our knowledge of different domains reasonably successfully when in reality they are each described in innumerable individually coherent ways but which, when brought together, do not amount to a unified and coherent world theory.

To understand and resolve this issue we might examine the very interesting distinction that arises between communication and computation when the computer is conceived as an open system in communication with the environment; so, in the case of biological computing where, crucially, the boundary is dynamic. For many of the authors in this volume biology is at the centre of an 'information revolution'. The new fields of, for example, genomics, proteomics, and bioinformatics are capturing both enormous investment and a great deal of the lime-light; and at the same time, they are producing copious realms of specialized biomolecular data. Paradoxically, not much theoretical effort is being devoted to the elaboration of an integrated 'informational perspective' of the living cell, or to make sense of the whole 'informational architectures' withstanding the multicelullar complexity evolved by eukaryots – the nascent integrative field of systems biology is mostly concerned with empirically-oriented problems.

In this regard, the Marijuán & Moral essay provides an interesting theoretical starting-point. It is based on a taxonomy of the molecular recognition events occurring in the cell, progressively advancing towards a unitary contemplation of the informational organization of life. Along the way, one can make sense of the overall convergence of informational architectures and signaling systems upon an organismic cycle decomposable in multiple cell-cycles. Of particular interest in their approach are the taxonomy of molecular recognition events, the stochastic nature of enzyme function, the combination of 'functional addresses' and 'secondary address' in a type of von Neumann computing scheme in the eukaryotic DNA, the functioning of the Cellular Signaling System, and their discussion of how the cell builds 'meaning' for the different combinations of received signals. Ultimately they argue that the organization of life, that which distinguishes animate from inanimate matter, would not evolve only on the existence of a genetic 'code', or on a series of nested 'codes', but on a series of evolutionary inventions of an 'informational' nature, that is, related to combinatoric arrangements of molecular recognition events.

Brier's contribution to this volume is written from the perspective of biosemiotics – the study of signs, communication and information in living systems. It is semiosis which raises living systems above the purely physical, chemical and, even, informational explanations, and, with Sebeok's work (*Viz.* Sebeok 1976, 1989) biosemiotics has extended to a description of semiosis between body cells and within the cells (*endosemiotics*) (Uexküll et al. 1993). Peirce’s semiotics deals with non-intentional bodily and natural signs, but un-interpreted 'natural' and technological objects are merely protosemiotic, displaying only 'Secondness' in Peircean terminology. For Brier the big question is whether a biosemiotic Peircian framework can encompass these protosemiotic objects. Following on from this he asks whether any of the human signs that computers manipulate can be said to be signs for the computer *per se*; in other words, why is it that we focus only on humans in language, culture and society as producers of meaning? In response he offers a cybersemiotics which embraces the evolutionary view of system science, cybernetics, and information science arguing that semiosis is "immanent in the universe, manifesting itself clearly in living systems and becoming emancipated and self-organized in social systems".

Carsetti's essay is also concerned with information patterns and meaning, but specifically with the emergence of meaning at the level of system-organisation. This meaning is necessarily connected to specific linguistic and logical operations, and to specific observational procedures. His emphasis is on a synergetic methodology – the formation and self-organiation of patterns and structures in 'open' systems. He examines Kanizsa's work in Gestalt theory and the problem of amodal completion – where we know perceptually that an object is occluded but we do not actually see it, arguing that a synergetic order-formation theory, like that used in non-linear dynamical systems, is the only one to offer a robust explanatory model of what we take the emerging pattern to mean. However, even this is not sufficient for the understanding we have of the temporally enduring conscious mind; for that we need a continuous interplay with the environment.

In the section on Cognitive Science and Philosophy we focus on consciousness, in juxtaposition with our other themes. Århem's paper addresses the question of consciousness at two different levels: on evolutionary level in neuroanatomical features critical for sustaining consciousness and at cellular level by studying the neural impulse patterns characteristic of a conscious state. He examines Crick and Koch's phylogenetic approach to consciousness extending it into and beyond Edelman's, Cotterill's and Eccles' work respectively. The thrust of this part of his essay is that to understand the evolution of consciousness we should focus on the reptilian brain; the reptile-bird brain evolution transition reveals a greater continuity than was initially expected, and this makes the consciousness cut-off point at birds particularly problematic. In the second part of his paper he emphasises the importance of understanding anaesthesiology because, although there is still very little known about what happens at either the system or the molecular level when anaesthetics are administered, we do know that they switch a network of neurons from high-frequency firing to low-frequency firing. Århem concludes that if further analysis reveals what it is that is being inhibited, we might have some better indication of what is important as the underpinning for consciousness.

In his essay on consciousness Pylkkanen focuses on the temporal structure of conscious experience when, for example, we hear music. He weaves together Husserl's tri-partite analysis of time consciousness – the primal impression of now, the retention of just past, and the protention of what is to come – with

Bohm's implicate order framework to present a critique of van Gelder's dynamical model of auditory pattern recognition and to present a richer account of the perception of previously experienced elements in our awareness of temporally enduring objects. Crucial to his account is the notion of the 'enfoldment' of these retained or previously experienced elements; a conclusion that presents a resolution for Husserl's seeming paradox of experiencing the past and future in the present.

Contemporary philosophy of mind is for the most part inspired by the naturalist intuition that the mind is part of the natural universe. A majority of contemporary philosophers assume a physicalist concept of mind according to which mental phenomena derive from neurophysiological phenomena, if they are not neurophysiological *per se*. In their attempt to naturalise the concept of mind many cognitive scientists rely on computational and / or informational metaphors and tools. This is the direction taken by the remaining three papers in this section of the book; each deals with the notion of computationalism and, more or less, explicitly the computational theory of mind (CTM). Brattico's interest is in Fodor's continuing objection to the CTM as providing an account of global cognition, that is, the kind of logically deep, complex thinking that goes in to deciding, for example, whether you let your children have the puppy they keep asking for or, in Brattico's much neater, but still computationally complex, example, whether or not you should carry an umbrella. This kind of everyday thinking involves pragmatic reasoning, problem solving, abductive reasoning and imagination, and in coming to a decision there must be a trade-off between efficiency and reason. In Artificial Intelligence modelling this becomes the 'frame problem' – distinguishing between computationally relevant and computationally irrelevant facts or, as Fodor puts it: "Hamlet's problem: when to stop thinking" (1987, p.140) – and a trade-off between rationality and computational complexity must be reached. In his attempt to resolve the problem of the poor trade-off between rationality and efficiency in cognitive modelling Brattico adopts Bennett's complexity theoretical notion of *logical depth* (Bennett 1988), and concludes that the logical depth of the knowledge underpinning our everyday decision-making is the result of an evolutionary process, a process which has been enriched by the learning of lessons communicated to us from previous generations.

Milkowski's focus is on the triviality threat to computationalism, that is, the fact that simultaneously some deny that cognition involves computation, yet others claim that all physical processes are computational (pancomputationalism), possibly even algorithmic, and he argues that even if these claims are true, computationalism need not be rendered trivial. He begins by distinguishing two main forms of computationalism: (i) that cognitive processes could be described algorithmically, and (ii) that cognitive processes are algorithmic or computational; and by distinguishing three varieties of computationalism: (i) that cognitive processes can be simulated computationally, (ii) that cognitive processes can be realised computationally, and (iii) that cognitive processes are generated by overall computational processes. In doing this he establishes a multi-level model of cognition where a base level, on which higher levels are supervenient, might be conceived of algorithmically but the emergent computational levels might not or, at least, not in the same algorithms. The result is a taxonomy of possible computationalisms, some of which are weak and trivial, and some of which are robust and non-trivial.

Finally, in this section, Lappi takes up the challenge to computational neuroscience presented by Grush (2001): How do you distinguish between computation – understood as computational processing of semantic information – and any other complex causal process that is governed by a computationally tractable rule? Grush's analysis of the problem is given in terms of a-semantics ( isomorphism between the causal neural processes and some abstract algorithm) and e-semantics (isomorphism between the causal neural process and the physical causal processes within the environment), and he claims that recent computational neuroscience treats computation and representation a-semantically and, thus, inadequately; what is needed is a more genuinely semantic notion of computation and representation in terms of e-semantics. Lappi agrees with Grush about the limitations of a-semantics but suggests, instead, a c-semantic construal of computation and representation, that is, one which makes essential reference to the cognitive mechanisms of the organism that represent a systematic and genuine contribution of the organism to semantic content and which provides not merely a reflection of structure in the environment.

From a concern with CTM, we move to applications in the domain of computational linguistics (CL) and, in the first instance, to ten Hacken's claim that CL should be conceived as an applied science. He begins by using the example of parsing to show that CL can be pursued as an empirical science, as an applied science or as mere technology. However, since applied sciences are problem-oriented, they have both a practical orientation and an explanatory focus; it is this explanatory focus, he argues, that facilitates the expression of well-formed problems in terms of their *identification*, *evaluation*, *decomposition*, and appropriate *knowledge selection*.

Knuuttila also makes, as she herself says, a "short philosophical visit into the world of parsing" but to her the interesting question is an ontological one about the manner in which we tend to treat language technological models, like the Constraint Grammar Parser, as abstract, theoretical representations of the world rather than as epistemic artefacts. In arguing that models should be treated more robustly and not simply as substitutes for the 'real' thing, Knuuttila emphasises the artificiality, workability and experimentability of scientific models, and suggests that in treating them less like Platonic ideals we might begin to rethink our approach to science more generally.

According to Hirst much of the success of recent computational linguistics (CL), natural language processing (NLP), and human language technologies (HLT) has been achieved by adapting the problem we are trying to solve. He presents three computational views of text-meaning: (i) objective text-meaning, (ii) authorial intent, and (iii) subjective text-meaning, and argues that if CL is to progress it must incorporate all three approaches and not just the first. Moreover, it must develop ways of dealing with the reader's purposiveness and the possibility of what the author might have said but didn't, not just with the familiar notions of pragmatics and implicature.

Dynamics, process, change, and agency are central concepts in the computationalist view. As a result of giving (human) agency and thus intentionality such a prominent role, computationalists must address the ethical aspects and consequences of that agency. Ethics must be an explicit constitutive element in the whole computationalist / informationalist epistemic enterprise.

The issues surrounding pancomputationalism, paninformationalism, biosemantics, cognition, ontology, logical pluralism, and technology are of paramount importance to us, yet they are academically liminal, awkwardly interdisciplinary with no clearly-defined disciplinary responsibility. Thus, no text that deals with these topics as they are currently perceived and conceived can fail to consider the relevant ethical and educational issues that accompany them.

To this end the final section of the book begins with an examination by Bynum of an ethics for the new millennium. Bynum's *Flourishing Ethics*, that draws its inspiration from biological systems, argues for ethical pluralism. In a context of globalization it is likely that this pluralism will become a commonplace reality for everybody and, for this reason alone, it deserves scholarly attention.

Since, it is argued, information ethics is presenting us with a revolution in ethical discourse we need a Copernican Revolution in ethical theory. This is exactly what Bynum's essay does; it introduces us to a nascent ethical theory in the field of Computer Ethics: *Flourishing Ethics*. It is a theory that will provide new tools for dealing with the 'policy vacuums' (Moor 1985) that are presented, almost daily, by the advent of new technologies and our current scientific understanding of life, human nature and the universe. It will also make it possible for us to revisit the shortcomings of already long-established ethical theories. Taking a lead from Aristotle, Bynum argues that excellence in information processing will produce excellence in action and, thus, human flourishing.

It is an invigorating ethics that stresses the role of human beings as caretakers rather than exploiters of others and our planet's resources. It encourages moral agents to see themselves not as independent from each other and the universe but as part of a continuum, the information, and possibly computation, continuum. "The shift in perspective advocated by Flourishing Ethics, then, brings human beings back into the fold with the rest of the universe. It views humans, like all other beings, as fellow participants in the creative unfolding of the cosmos – fellow travelers in the cosmic river of flowing information." (Bynum, this volume)

Riegler analyses the paradox of autonomy in which he argues quite forcefully that when an artificial, self-organising system has reached a state of proper autonomy, that is, where their goals are independent from our own, it will be impossible for us to interact with them. The consequences of this are, it is needless to say, serious and never more so than in the context of personal service robots – systems that are being developed to aid in the care of the elderly and those a suffering chronic illness. The development of increasingly *e*-autonomous agents and the autonomy paradox that this might present is something we must consider carefully, especially if we want their interaction with us to continue in our favour.

Trojer depicts the challenges a technical university, with an explicit profile of applied Information and Communication Technology (ICT), encounters in a region with strong development and when the cooperation with public and private partners outside the university becomes a necessary and predominant reality. It is a situation that calls for knowledge and information transformation processes, and an epistemological openness among people active at the university is a prerequisite for successful functional cooperation. The main questions concern resources for staying confident, future-oriented and innovative as an ICT researcher and a member of the academic teaching staff. Trojer makes reference to her five year development experience, involving student recruitment, research and campus building, and the securing of resources for the epistemological framework within which she has developed a feminist techno-science programme within a technical faculty. Her paper contributes to the epistemological pluralism and multiculturalism of the book, alongside Baynum's ethical pluralism and Allo's logical pluralism.

The final essay in this volume, and another which addresses the issue of computer ethics, has been contributed by Brey. In his paper Brey states that "Computer ethics is a major new field of study that addresses ethical issues in the use, development and management of information technology, as well as in the formulation of general societal policies regarding the regulation of information technology in society". Universities are beginning to develop the teaching of computer ethics as a subject and, he maintains, they must concentrate on two central areas: (i) *computer ethics policies*, and (ii) *computer ethics education*. Brey presents an outline for a field of study which he refers to as social and humanistic studies of computing (SHC) and contrasts it with the applied study of societal aspects of computing (ASC). Both must have a place in the university curriculum if they are to generate awareness of the complex issues we now confront.

We have brought this collection of essays together in an attempt to establish the interconnecting themes within a pancomputational and paninformational, process-oriented conception of the universe. The interdisciplinary nature of the essays can occasionally mean that they exist in the liminal regions of two or more fields of inquiry, the result of which can be that they are overlooked in a discipline-orientated academy. What we have established, in bringing them together, is a perspective and a series of perspectives from, on, and within a burgeoning domain where no single discipline is pre-eminent.

That these particular essays have been brought together does not imply that they provide a set of answers, or even necessarily that there are answers, but in your interaction with them we anticipate that a further series of dialogues will be established. Thus, we offer them both individually and in their compilation as "new metaphors, new starting points". [Winograd 1997]

## References

Beall JC & Restall G, 2000. Logical Pluralism, Australasian Journal of Philosophy, 78, pp 475–493

Bennett, C. H. 1988. 'Logical depth and physical complexity', in The Universal Turing Machine; A Half-Century Survey, 227—257. Edited by R. Herken. Oxford: Oxford University Press

Dodig-Crnkovic G. 2006. Investigations into Information Semantics and Ethics of Computing, Mälardalen University Press, forthcoming

Floridi, L. 2004. 'Open Problems in the Philosophy of Information', Metaphilosophy, 35.4

Fodor, J.A. 1987. 'Modules, frames, fridgeons, sleeping dogs and the music of the spheres'. In The robot's dilemma: The frame problem in artificial intelligence, ed. Z. Pylyshyn. Norwood, NJ: Ablex.

Grush, R. 2001. 'The Semantic Challenge to Computational Neuroscience', in Machamer, P. Grush, R. & McLaughlin, P. (Eds.), Theory and Method in the Neurosciences, Pittsburgh, PA: University of Pittsburgh Press.

Hilbert, D, 1926. 'Über das Unendliche'. Mathematische Annalen, 95: 161—90. Translated as "On the infinite" in van Heijenoort, From Frege to Gödel: A source book in mathematical logic, 1879-1931, Harvard University Press

Kafatos, M. & Nadeau, R. 1999. The Conscious Universe: Parts and Wholes in Physical Reality, Springer, 2nd Edition

Kant, I. 1929. The Critique of Pure Reason, trans. Norman Kemp Smith, Macmillan Press (A edition 1781 + B edition 1787)

Lloyd, S. 2006. Programming the Universe: A Quantum Computer Scientist Takes on the Cosmos, Jonathan Cape

Moor, J. H. 1985. "What Is Computer Ethics?", in Terrell Ward Bynum, ed., Computers and Ethics, Blackwell, pp. 266-75 (first published in Metaphilosophy October 1985)

Sebeok, T. 1976. Contributions to the Doctrine of Signs. Bloomington: Indiana University Press

—. 1989 The Sign & Its Masters. Sources in Semiotics VIII. New York: University Press of America.

Uexküll, Thure von, Geigges, W., Herrmann J. M. 1993. 'Endosemiosis', Semiotica 96 (1/2), pp 5-51

Winograd, T. 1997. The design of interaction. Beyond Calculation: The next 50 years of computing. Copernicus, pp 149-161

Zuse, K. 1967. Rechnender Raum, Elektronische Datenverarbeitung, vol. 8, pp 336-344

# PART I:
# INFORMATION

# CHAPTER ONE

# EPISTEMOLOGY AS INFORMATION THEORY: FROM LEIBNIZ TO Ω

GREGORY CHAITIN

## Abstract

In 1686 in his *Discours de métaphysique,* Leibniz points out that if an arbitrarily complex theory is permitted then the notion of "theory" becomes vacuous because there is always a theory. This idea is developed in the modern theory of algorithmic information, which deals with the size of computer programs and provides a new view of Gödel's work on incompleteness and Turing's work on uncomputability. Of particular interest is the halting probability Ω, whose bits are irreducible, i.e., maximally unknowable mathematical facts. More generally, these ideas constitute a kind of "digital philosophy" related to recent attempts of Edward Fredkin, Stephen Wolfram and others to view the world as a giant computer. There are also connections with recent "digital physics" speculations that the universe might actually be discrete, not continuous. This *système du monde* is presented as a coherent whole in my book *Meta Math!,* which will be published this fall.

## Introduction

I am happy to be here with you enjoying the delicate Scandinavian summer; if we were a little farther north there wouldn't be any darkness at all. And I am especially delighted to be here delivering the Alan Turing Lecture. Turing's famous 1936 paper is an intellectual milestone that seems larger and more important with every passing year.

[For Turing's original paper, with commentary, see Copeland's The Essential Turing.]

People are not merely content to enjoy the beautiful summers in the far north, they also want and need **to understand**, and so they create myths. In this part of the world those myths involve Thor and Odin and the other Norse gods.

In this talk, I'm going to present another myth, what the French call a *système du monde,* a system of the world, a speculative metaphysics based on information and the computer.

[One reader's reaction (GDC): "Grand unified theories may be like myths, but surely there is a difference between scientific theory and any other narrative?" I would argue that a scientific narrative is more successful than the Norse myths because it explains what it explains more precisely and without having to postulate new gods all the time, i.e., it's a better "compression" (which will be my main point in this lecture; that's how you measure how successful a theory is).]

The previous century had logical positivism and all that emphasis on the philosophy of language, and completely shunned speculative metaphysics, but a number of us think that it is time to start again. There is an emerging digital philosophy and digital physics, a new metaphysics associated with names like Edward Fredkin and Stephen Wolfram and a handful of like-minded individuals, among whom I include myself. As far as I know the terms "digital philosophy" and "digital physics" were actually invented by Fredkin, and he has a large website with his papers and a draft of a book about this. Stephen Wolfram attracted a great deal of attention to the movement and stirred up quite a bit of controversy with his very large and idiosyncratic book on *A New Kind of Science.*

And I have my own book on the subject, in which I've attempted to wrap everything I know and care about into a single package. It's a small book, and amazingly enough it's going to be published by a major New York publisher a few months from now. This talk will be an overview of my book, which presents my own personal version of "digital philosophy," since each of us who works in this area has a different vision of this tentative, emerging world view. My book is called *Meta Math!*, which may not seem like a serious title, but it's actually a book intended for my professional colleagues as well as for the general public, the high-level, intellectual, thinking public.

"Digital philosophy" is actually a neo-Pythagorean vision of the world, it's just a new version of that. According to Pythagoras, all is number — and by number he means the positive integers, 1, 2, 3, ... — and God is a mathematician. "Digital philosophy" updates this as follows: Now everything is made out of 0/1 bits, everything is digital software, and God is a computer programmer, not a mathematician! It will be interesting to see how well this vision of the world succeeds, and just how much of our experience and theorizing can be included or shoe-horned within this new viewpoint.

[Of course, a system of the world can only work by omitting everything that doesn't fit within its vision. The question is how much will fail to fit, and conversely, how many things will this vision be able to help us to understand. Remember, if one is wearing rose colored glasses, everything seems pink. And as Picasso said, theories are lies that help us to see the truth. No theory is perfect, and it will be interesting to see how far this digital vision of the world will be able to go.]

Let me return now to Turing's famous 1936 paper. This paper is usually remembered for inventing the programmable digital computer via a mathematical model, the Turing machine, and for discovering the extremely fundamental halting problem. Actually Turing's paper is called "On computable numbers, with an application to the *Entscheidungsproblem,*" and by computable numbers Turing means "real" numbers, numbers like $e$ or $\pi = 3.1415926...$ that are measured with infinite precision, and that can be computed with arbitrarily high precision, digit by digit without ever stopping, on a computer.

Why do I think that Turing's paper "On computable numbers" is so important? Well, in my opinion it's a paper on epistemology, because we only understand something if we can program it, as I will explain in more detail later. And it's a paper on physics, because what we can actually compute depends on the laws of physics in our particular universe and distinguishes it from other possible universes. And it's a paper on ontology, because it shows that some real numbers are **uncomputable**, which I shall argue calls into question their very existence, their mathematical and physical existence.

[You might exclaim (GDC), "You can't be saying that before Turing and the computer no one understood anything; that can't be right!" My response to this is that before Turing (and my theory) people could understand things, but **they couldn't measure how well** they understood them. Now you can measure that, in terms of the degree of compression that is achieved. I will explain this later at the beginning of the section on computer epistemology. Furthermore, programming something forces you to understand it better, it forces you to really understand it, since you are explaining it **to a machine**. That's sort of what happens when a student or a small child asks you what at first you take to be a stupid question, and then you realize that this question has in fact done you the favor of forcing you to formulate your ideas more clearly and perhaps even question some of your tacit assumptions.]

To show how strange uncomputable real numbers can be, let me give a particularly illuminating example of one, which actually preceded Turing's 1936 paper. It's a very strange number that was invented in a 1927 paper by the French mathematician Emile Borel. Borel's number is sort of an anticipation, a partial anticipation, of Turing's 1936 paper, but that's only something that one can realize in retrospect. Borel presages Turing, which does not in any way lessen Turing's important contribution that so dramatically and sharply clarified all these vague ideas.

[I learnt of Borel's number by reading Tasic's *Mathematics and the Roots of Postmodern Thought,* which also deals with many of the issues discussed here.]

Borel was interested in "constructive" mathematics, in what you can actually compute we would say nowadays. And he came up with an extremely strange non-constructive real number. You list all possible yes/no questions in French in an immense, an infinite list of all possibilities. This will be what mathematicians call a denumerable or a countable infinity of questions, because it can be put into a one-to-one correspondence with the list of positive integers 1, 2, 3, ... In other words, there will be a first question, a second question, a third question, and in general an *N*th question.

You can imagine all the possible questions to be ordered by size, and within questions of the same size, in alphabetical order. More precisely, you consider all possible strings, all possible finite sequences of symbols in the French alphabet, including the blank so that you get words, and the period so that you have sentences. And you imagine filtering out all the garbage and being left only with grammatical yes/no questions in French. Later I will tell you in more detail how to actually do this. Anyway, for now **imagine** doing this, and so there will be a first question, a second question, an *N*th question.

And the *N*th digit or the *N*th bit after the decimal point of Borel's number answers the *N*th question: It will be a 0 if the answer is no, and it'll be a 1 if the answer is yes. So the binary expansion of Borel's number contains the answer to every possible yes/no question! It's like having an oracle, a Delphic oracle that will answer every yes/no question!

How is this possible?! Well, according to Borel, it isn't really possible, this can't be, it's totally unbelievable. This number is only a mathematical fantasy, it's not for real, it cannot claim a legitimate place in our ontology. Later I'll show you a modern version of Borel's number, my halting probability Ω. And I'll tell you why some contemporary physicists, real physicists, not mavericks, are moving in the direction of digital physics.

[Actually, to make Borel's number as real as possible, you have to avoid the problem of filtering out all the yes/no questions. And you have to use decimal digits, you can't use binary digits. You number all the possible finite strings of French symbols including blanks and periods, which is quite easy to do using a computer. Then the *N*th digit of Borel's number is 0 if the *N*th string of characters in French is ungrammatical and not proper French, it's 1 if it's grammatical, but not a yes/no question, it's 2 if it's a yes/no question that cannot be answered (e.g., "Is the answer to this question "no"?"), it's 3 if the answer is no, and it's 4 if the answer is yes.]

*Geometrically* a real number is the most straightforward thing in the world, it's just a point on a line. That's quite natural and intuitive. But *arithmetically*, that's another matter. The situation is quite different. From an arithmetical point of view reals are extremely problematical, they are fraught with difficulties!

Before discussing my Ω number, I want to return to the fundamental question of what does it mean to understand. How do we explain or comprehend something? What is a theory? How can we tell whether or not it's a successful theory? How can we measure how successful it is? Well, using the ideas of information and computation, that's not difficult to do, and the central idea can even be traced back to Leibniz's 1686 *Discours de métaphysique.*

## Computer Epistemology: What is a mathematical or scientific theory? How can we judge whether it works or not?

In Sections V and VI of his *Discourse on Metaphysics,* Leibniz asserts that God simultaneously maximizes the variety, diversity and richness of the world, and minimizes the conceptual complexity of the set of ideas that determine the world. And he points out that for any finite set of points there is always a mathematical equation that goes through them, in other words, a law that determines their positions. But if the points are chosen at random, that equation will be extremely complex.

This theme is taken up again in 1932 by Hermann Weyl in his book *The Open World* consisting of three lectures he gave at Yale University on the metaphysics of modern science. Weyl formulates Leibniz's crucial idea in the following extremely dramatic fashion: If one permits arbitrarily complex laws, then the concept of law becomes vacuous, because there is always a law! Then Weyl asks, how can we make more precise the distinction between mathematical simplicity and mathematical complexity? It seems to be very hard to do that. How can we measure this important parameter, without which it is impossible to distinguish between a successful theory and one that is completely unsuccessful?

This problem is taken up and I think satisfactorily resolved in the new mathematical theory I call *algorithmic information theory.* The epistemological model that is central to this theory is that a scientific or mathematical theory is a computer program for calculating the facts, and the smaller the program, the better. The complexity of your theory, of your law, is measured in bits of software:

program (bit string) —> **Computer —>** output (bit string)

theory —> **Computer —>** mathematical or scientific facts