Algorithmic Social Sciences Research Unit

ASSRU

Department of Economics
University of Trento
Via Inama 5
381 22 Trento Italy

Discussion Paper Series

1 – 2011/II

THE COMPUTABILITY – THEORETIC CONTENT OF EMERGENCE♠

S. Barry Cooper

May 2011

♠ Text of the talk given at the ASSRU/Department of Economics Seminar, University of Trento, 11 May, 2011. Professor Cooper is a *Founding Honorary Associate* of ASSRU.

The Computability-Theoretic Content of Emergence

S. Barry Cooper
Department of Pure Mathematics
University of Leeds, Leeds LS2 9JT, U.K.

Abstract

In dealing with emergent phenomena, a common task is to identify useful descriptions of them in terms of the underlying atomic processes, and to extract enough computational content from these descriptions to enable predictions to be made. Generally, the underlying atomic processes are quite well understood, and (with important exceptions) captured by mathematics from which it is relatively easy to extract algorithmic content.

A widespread view is that the difficulty in describing transitions from algorithmic activity to the emergence associated with chaotic situations is a simple case of complexity outstripping computational resources and human ingenuity. Or, on the other hand, that phenomena transcending the standard Turing model of computation, if they exist, must necessarily lie outside the domain of classical computability theory.

In this article we suggest that much of the current confusion arises from conceptual gaps and the lack of a suitably fundamental model within which to situate emergence. We examine the potential for placing emergent relations in a familiar context based on Turing's 1939 model for interactive computation over structures described in terms of reals. The explanatory power of this model is explored, formalising informal descriptions in terms of mathematical definability and invariance, and relating a range of basic scientific puzzles to results and intractable problems in computability theory.

1 Computability and Emergence

Since the time of Turing, computability as a concept has hardened, become hedged around by its impressive technical development and its history, until its role from almost any viewpoint has become tangential to the very real mysteries of how one models the real universe. This turn of events has had an air of inevitability, in that even Turing, with his remarkable ability for clarifying concepts and basic questions, was unable to fully import his concerns about the nature of computability into the burgeoning formal framework of recursion theory. And many of those who took up the technical development of the subject not only lacked Turing's vision, but became diverted by the pure excitement and mathematical beauty of the new academic field. Thomas Kuhn's 'normal science' contains its own excitements and its minor paradigm shifts, as well as delivering safe research careers.

From the point of view of the logician, recursion theory concerns on the one hand a *computable* universe whose theory derives its significance from computer-scientific concerns, with a technical content owing only a very basic and vestigial debt to its logical origins. And on the other hand, exhibits an arcane preoccupation with the development of a theory of *incomputability*, for which its practitioners have no explanation or evidence for its existence in the material world. One may be uneasy about the public criticisms by Martin Davis, Stephen Simpson, and others (see [12]), but their views are widely respected.

This leaves many, with eyes wide enough open to see the accumulated evidence of real-world misbehaviour, looking elsewhere for models. Presented with phenomena with seemingly no hope of ever being reduced to a simple classical computational model, the natural alternative has been to develop models with direct links to quite particular instances of apparent incomputability in a physical setting. Much of this work, giving rise to a wide range of so-called 'new computational paradigms', has taken on a distinctly ad hoc aspect. Even though the theoretical underpinnings of this newness are absent — even the standard model of quantum computation is not free from continued scrutiny — the delivery of computational outcomes sufficiently separated from the model's real-world template is taken as a pointer to useful applications.

One can highlight three key challenges to a reductionist view of the computational content of the universe, and to the explanational potentialities of the computability framework. All three are familiar to the informed non-specialist, are strikingly hard for the specialist to deal with, and are associated with controversies, speculations, and a missing clarity which suggests a corresponding missing conceptual ingredient. Quantum phenomena, and the human brain, present the two most unavoidable challenges to the reductionist agenda. There are other relatively specific examples, such as the puzzle of the origins of life. But these are less dramatic, and less in the public domain. The third challenge — *emergence* — is at first sight less obviously disturbing, but is more prevalent, more *protean* in its manifestations, more theoretically deconstructable, and — ultimately — more likely to give rise to a basic theoretical model of wide application. And potentially of wide enough relevance to throw light on the two first and more immediate challenges to our understanding of the world.

Emergence lies at the core of a number of controversies in science, often used in a descriptive and speculative way to challenge more mechanistic and reductive attempts to interpret the universe. Out of this dichotomy arises a less-than-illuminating polarisation into a relative faithfulness to the simpler Laplacian constructs of the scientific age, and a contemporary counter-culture insistent on the essential mystery and predominance of emergent phenomena. The purpose of this article is to point to some sort of reconciliation, mediated by classical computability concepts going back to Turing — the unifying personality both in his overall concerns with computability, and in his breadth of interests, taking in his seminal work on emergence, in the form of his work on morphogenesis, and specifically phyllotaxis.

2 What is Emergence?

The term *emergence* is increasingly used in all sorts of contexts, often to describe any situation in which there appears to be a breakdown in reductionist explanation, or where there appears to be a global rather than purely local causal dynamic at work. This is how Stuart Kauffman [26] argues in his recent book on *Reinventing the Sacred: A New View of Science, Reason and Religion* (p.281):

> We are beyond reductionism: life, agency, meaning, value, and even consciousness and morality almost certainly arose naturally, and the evolution of the biosphere, economy, and human culture are stunningly creative often in ways that cannot be foretold, indeed in ways that appear to be partially lawless. The latter challenge to current science is radical. It runs starkly counter to almost four hundred years of belief that natural laws will be sufficient to explain what is real anywhere in the universe, a view I have called the Galilean spell. *The new view of emergence and ceaseless creativity partially beyond natural law is a truly new scientific worldview in which science itself has limits.* [My emphasis.]

If one is going to give emergence such a key role in restructuring the Laplacian model of science, and to come up with a suitably basic explanatory model, one needs to be more clear about what are the defining characteristics of emergent phenomena. Ronald, Sipper and Capcarrère [37] draw a parallel with the development of the Turing Test for intelligent machines, and use Turing's observer-based approach to formulate an *emergence test*. They comment that "overly facile use of the term emergence has made it controversial", and quote Arkin [2, p.105]:

> Emergence is often invoked in an almost mystical sense regarding the capabilities of behavior-based systems. Emergent behavior implies a holistic capability where the sum is considerably greater than its parts. It is true that what occurs in a behavior-based system is often a surprise to the system's designer, but does the surprise come because of a shortcoming of the analysis of the constituent behavioral building blocks and their coordination, or because of something else?

Ronald, Sipper and Capcarrère's emergence test "centers on an observer's avowed incapacity (amazement) to reconcile his perception of an experiment in terms of a global world view with his awareness of the atomic nature of the elementary interactions". As well as an observer, there is a 'designer' in the picture, whose existence is used to assist the description of certain qualifying features of the atomic interactions of the system to be tested. The test is comprised of three criteria:

1. *Design:* The system has been constructed by the designer, by describing *local* elementary interactions between components (e.g., artificial creatures and elements of the environment) in a language $\mathcal{L}_1$.

2. *Observation:* The observer is *fully aware* of the design, but describes *global* behaviors and properties of the running system, over a period of time, using a language $\mathcal{L}_2$.

3. *Surprise:* The language of design $\mathcal{L}_1$ and the language of observation $\mathcal{L}_2$ are distinct, and the causal link between the elementary interactions programmed in $\mathcal{L}_1$ and the behaviors observed in $\mathcal{L}_2$ is *non-obvious* to the observer — who therefore experiences surprise. In other words, there is a cognitive dissonance between the observer's mental image of the system's design stated in $\mathcal{L}_1$ and his contemporaneous observation of the system's behavior stated in $\mathcal{L}_2$.

Ronald, Sipper and Capcarrère elaborate on this third condition to eliminate evanescent instances of surprise. Notice that one can apply versions of these criteria to a wide range of situations in which one is effectively capable of 'looking over the shoulder' of a putative designer — say one in which the local science is handed down to us by Nature, and is though to be well-understood, e.g., self-contained systems implementing Newtonian laws. The early history of chaos theory is replete with examples exhibiting the right quality of surprisingness, nicely communicated by the term 'strange attractor' coined [38] by David Ruelle and Floris Takens in 1971.

Of course, there is now quite a long history (see for example [1]) aimed at describing and improving our understanding of emergence, and as time goes on the observer 'surprise' criterion may not be as robust as the corresponding element of the Turing test. Turing himself played an innovative role in developing demystifying mathematics related to morphogenesis, and more particularly phyllotaxis, both in his seminal published paper [44] on the mathematical theory of biological pattern formation, and in his more opaque and incomplete writings contained in the posthumous collected works [45].

What is important though is not just the demystifying role of descriptions of emergent phenomena, but the *representational* functionality they point to. It is this latter aspect that takes us beyond emergence to a view of complexity in Nature in which emergence plays a key *inductive* role. And it is the first two of Ronald, Sipper and Capcarrère's conditions which make us look for something else within particular highly complex situations in which emergence clearly plays a role, though not a definitive one. These first two conditions also point to the a route to isolating the computational content of aspects of the physical universe which appear on the one hand to transcend standard computability-theoretic frameworks, and on the other entice reductionist explanations of increasing implausibility.

3 Representations, Recursions, Memetic Transmission

In [9] we considered the computational content of features of the real world, and more particularly, of developing computational practice. We looked at instances in which there appeared to be a fairly basic transgression of the 'Turing barrier' (defined by the limit of what is computable by an ideal computer as captured theoretically by a universal Turing machine), and more complex examples such as human intelligence and quantum uncertainty. In the former case one finds the emergence test broadly applicable, and in so doing can get a more informative theoretical grasp of what emergence is as a computational process.

For instance, going back to the influential 1988 paper of Paul Smolensky in *Behavioral and Brain Sciences*, we find [39, p.3] him examining a model qualifying under criteria one and two of the emergence test, along with an indication of an outcome which is surprising, judged according to computability-theoretic expectations:

> There is a reasonable chance that connectionist models will lead to the development of new somewhat-general-purpose self-programming, massively parallel analog computers, and a new theory of analog parallel computation: they may possibly even challenge the strong construal of Church's Thesis as the claim that the class of well-defined computations is exhausted by those of Turing machines.

Computational parallelism is an obviously important aspect of connectionist models and many others, but one needs to be careful about claiming that this is not simulated by a Turing machine. As is well-known (see, for example, David Deutsch [19, p.210]), the parallelism delivered by the standard model of quantum computation can be explained within the classical sequential model. A key ingredient, the addition of which does seem to stretch the classical Turing model, is that of internal connectivity. Goldin and Wegner [23] quote from Robin Milner's 1991 Turing Award lecture [30, p.80]:

> Through the seventies, I became convinced that a theory of concurrency and interaction requires a new conceptual framework, not just a refinement of what we find natural for sequential computing.

At the same time, parallelism and interactivity do seem to be basic features of situations exhibiting emergence.

Another idea which runs through a number of hypercomputational proposals, including Copeland's [15] rediscovery of oracle Turing machines, is that of adding *contextual* interactions. But as Davis has argued effectively, there is plenty of scope to widen the definition of what is 'internal' to a given system to bring a proposed new computational paradigm based on inadequately sourced oracles back into the classical fold.

But in [24], for instance, Goldin and Wegner are not just talking about parallelism and internal interactivity. And the inherent vagueness of examples they quote both stretch the mathematical analysis, and the reductionist agenda which feeds on that, to its limits:

> One example of a problem that is not algorithmic is the following instruction from a recipe [31]: 'toss lightly until the mixture is crumbly.' This problem is not algorithmic because it is impossible for a computer to know how long to mix: this may depend on conditions such as humidity that cannot be predicted with certainty ahead of time. In the function-based mathematical worldview, all inputs must be specified at the start of the computation, preventing the kind of feedback that would be necessary to determine when it's time to stop mixing.

But such interactions, such as those involving physical oracles as in [3], appear to take us beyond an analysis directly relevant to the computational ingredients of emergence as a basic computational phenomenon, and towards the more hybrid computational environments presaged at the end of the previous section.

A computational context which is both widely suspected of transcending the standard Turing model, and of whose inner workings we have a high level of detailed knowledge, is the human brain. And although we do know a great deal about the way the human brain works, it clearly fails to satisfy the first two conditions of the emergence test.

Part of the brain's potential for enrichment of our modelling of the computationally complex lies in the way it seems to successfully deal with the sort of imaging of the real world we would dearly like our computing machines to perform. More important, the brain shows the capacity to perform re-presentations of mental imaging to enable recursive development of complex conceptual structures. At the same time, new techniques for relating structural and functional features of the brain, for example, using positron emission scan (PET) or a functional magnetic resonance imaging scan (fMRI), bring us much closer to obtaining useful models.

As we noted in [9], connectionist models of computation based on the workings of the human brain have developed in sophistication since Turing's [43] discussion of 'unorganised machines' (cf. Jack Copeland and Diane Proudfoot's article [16] *On Alan Turing's Anticipation of Connectionism*), and McCulloch and Pitts' initial paper [32] on neural nets. But despite the growth of computational neuroscience as an active research area, putting together ingredients from both artificial neural networks and neurophysiology, something does seem to be missing. As Rodney Brooks [5] says "neither AI nor Alife has produced artifacts that could be confused with a living organism for more than an instant." Or as Steven Pinker puts it: "...neural networks alone cannot do the job", going on to describe [34, p.124] "a kind of mental fecundity called recursion":

> We humans can take an entire proposition and give it a role in some larger proposition. Then we can take the larger proposition and embed it in a still-larger one. Not only did the baby eat the slug, but the father saw the baby eat the slug, and I wonder whether the father saw the baby eat the slug, the father knows that I wonder whether he saw the baby eat the slug, and I can guess that the father knows that I wonder whether he saw the baby eat the slug, and so on.

We are good at devising computational models capable of imaging, and of going some way to emulate how the brain comes up with neural patterns representing quite complex formations. But the mechanisms the brain uses to represent such patterns and relate them in complex ways is more elusive. What makes the sort of recursion Stephen Pinker has in mind so difficult to get to grips with at the structural level, is that it seems wound up with the puzzle of consciousness and its relationship to emotions and feelings. Antonio Damasio [17, p.169] describes the hierarchical development of a particular instance of consciousness within the brain (or, rather, 'organism'), interacting with some external object:

> ... both organism and object are mapped as neural patterns, in first-order maps; all of these neural patterns can become images. ... The sensorimotor maps pertaining to the object cause changes in the maps pertaining to the organism. ... [These] changes ... can be re-represented in yet other maps (second-order maps) which thus represent the relationship of object and organism. ... The neural patterns transiently formed in second-order maps can become mental images, no less so than the neural patterns in first-order maps.

What is important here is the re-representation of neural patterns formed across some region of the brain, in such a way that they can have a computational relevance in forming new patterns. This is where the clear demarcation between computation and computational effect becomes blurred. The key conception is of computational loops incorporating these 'second-order' aspects of the computation itself. Building on this one can derive a plausible schematic picture of of the global workings of the brain.

Considering how complex a structure the human brain is, it is surprising one does not find more features needing reflecting in any basic computational model based on it. However, a thorough trawl through the literature, and one's own experiences, fails to bring to light anything that might be held up as computational principle transcending in a fundamental way what we have already identified. The key ingredients we expect in a model are imaging, parallelism, interconnectivity, and a counterpart to the second-order recursions pointed to above.

Mathematically, the imaging appears to be dependent on the parallelism and interconnectivity. This is what connectionist models are strong on. The recursions are not so easy to model, though. Looked at logically, one has representations of complex patternings of neural events underlying which there is no clear local mechanism, but for which one would expect a description in terms of the structures pertaining. Looked at physically, such descriptions appear to emerge, and be associated with (but not exclusively) the sort of non-linear mathematics governing the emergence of new relations from chaotic environments. This leads us to turn the picture of re-representations of mental imaging as a describable mapping on its head, and think (see [8]) in terms of descriptions in terms of a structure *defining*, and hence determining, the mental re-representations.

Looking at this more closely, what seems to be happening is that the brain stores away not just the image, but a route to accessing that image as a whole. This is what people who specialise in memorising very long numbers seem to display — rather than attempting to go directly into the detailed memory of a given number, they use simple representational tricks to call the entire number up. Here is how Damasio summarises the process (and the quotation from [17, p.170] is worth giving in full):

> As the brain forms images of an object — such as a face, a melody, a toothache, the memory of an event — and as the images of the object *affect* the state of the organism, yet another level of brain structure creates a swift nonverbal account of the events that are taking place in the varied brain regions activated as a consequence of the object-organism interaction. The mapping of the object-related consequences occurs in first-order neural maps representing the proto-self and object; the account of the *causal relationshi*p between object and organism can only be captured in second-order neural maps. . . . one might say that the swift, second-order nonverbal account narrates a story: *that of the organism caught in the act of representing its own changing state as it goes about representing something else.*

So what is going on here, and how can one make sense of this in a fundamental enough way to apply to it computability-theoretic analysis? Let us describe what seems to be the key idea in abstract terms, and then reinforce this powerful conceptual lever via something more familiar, but with new eyes.

What we first looked at, in a fairly schematic way, is a particular physical system whose constituents are governed by perfectly well-understood basic rules. These rules are usually *algorithmic*, in that they can be described in terms of functions simulatable on a computer, and their simplest consequences are mathematically predictable. But although the global behaviour of the system is *determined* by this algorithmic content, it may not itself be recognisably algorithmic. We certainly encounter this in the mathematics, which may be *nonlinear* and not yield the exact solutions needed to retain predictive control of the system. We may be able to come up with a perfectly precise *description* of the system's development which does not have the predictive — or algorithmic — ramifications the atomic rules would lead us to expect.

If one is just looking for a broad understanding of the system, or for a prediction of selected characteristics, the description may be sufficient. Otherwise, one is faced with the practical problem of extracting some hidden algorithmic content, perhaps via useful approximations, special cases, or computer simulations. Geroch and Hartle [22] discuss this problem in their 1986 paper, in which they suggest that "quantum gravity does seem to be a serious candidate for a physical theory for whose application there is no algorithm."

For the logician, this is a familiar scenario, for whom something describable in a structure is said to be *definable*. The difference between computability and definability is well-known. For example, if you go to any basic computability text (e.g., Cooper [7]) you will find in the *arithmetical hierarchy* a usable metaphor for what is happening here. What the arithmetical hierarchy encapsulates is the smallness of the computable world in relation to what we can describe. And Post's Theorem [35] shows us how language can be used to progressively describe increasingly incomputable objects and phenomena within computable structures. An analysis of lower levels of the hierarchy even gives us a clue to the formal role of computable approximations in constraining objects computably beyond our reach.

Now, the important thing to notice is that a description in some language can be viewed as being essentially *a code for an algorithm for reconstruction meaning from the real world within the human brain.* More precisely, a description conveys an epistemological algorithm which enables us to emulate emergent aspects, non-algorithmic, aspects of the world within the architecture of the brain. Key to this is the logical structure of the relevant word, sentence, or more extensive module of language. This, of course, is why certain ideas or human creations have *memetic content.* They come with a representation of, a recipe for, their mental recreation and simulation. The simulated phenomenon may be far from being algorithmic in its full manifestation, but the brain may be able to by-pass the computational barriers via an algorithmic device for activating and directing the brain's capacity for reproducing its own emergent features.

Of course, this process depends on humanly constructed language. But the universe has the capacity to handle descriptions, memetic content, and codings for algorithms which perform hugely sophisticated tasks, in a wide spectrum of situations, even though this may be via ad hoc emergent language of its own. Probably the most familiar example of this is the reproduction of various life forms via chromosomes and other genetic materials. A chromosome is a structured package of DNA and DNA-bound protein, involving genes, regulatory elements and other nucleotide sequences. Its coding functionality has algorithmic content, enabling the reproduction of complex aspects of the world — but this only within a context which is not obviously algorithmic, and which seems to ride upon undeniably emergent processes. Another example, involving the human brain, but not a particular language, is the process whereby experts in such tasks remember long seemingly random numbers. This is commonly achieved by algorithmically coding the details of the numbers into images simulable in the brain, the simulation itself being dependent upon higher order mental processes.

In order to associate a sufficiently basic model with such situations, which replaces the simple Laplacian determinism captured via Turing computability, one needs to look more closely at how science describes the world, and at the scientist's historic agenda. In particular, we will need to look at Turing's 1939 extension of his basic machine model of computation. The aim will be to go beyond an analysis of the computability-theoretic content of emergence, to that leading to a better understanding of the computational role of emergence in the wider context.

4 The Turing Model

Turing's extended [42] 1939 model, able to capture the algorithmic content of those structures which are presented in terms of real numbers can be seen in implicit form in Newton's *Principia*, published some 272 years earlier. Newton's work established a more intimate relationship between mathematics and science, and one which held the attention of Turing, in various guises, throughout his short life (see Hodges [25]). Just as the history of arithmetically-based algorithms, underlying many human activities, eventually gave rise to models of computation such as the Turing machine, so the oracle Turing machine schematically addresses the scientific focus on the extraction of predictions governing the form of computable relations over the reals. Whereas the inputting of data presents only time problems for the first model, the second model is designed to deal with possibly incomputable inputs, or at least inputs for which we do not have available an algorithmic presentation. One might reasonably assume that data originating from observation of the real world carries with it some level of computability, but we are yet to agree a mathematical model of physical computation which dispenses with the relativism of the oracle Turing machine. In fact, even as the derivation of recognisable incomputability in mathematics arises from quantification over algorithmic objects, so definability may play an essential role in fragmenting and structuring the computational content of the real world. The Turing model of computability over the natural numbers appears to many people to be a poor indicator of what to expect in science.

Typically, specialist computability theorists are loath to speculate about real-world significance for their work. Since the time of Turing, the theory of computability has taken on a Laputa-like[1] aspect in the eyes of many people, an arcane world disconnected from naturally arising information. Below, we look at Post's legacy of relating computability-theoretic concepts to intuitively immediate information content, and examine how that can be further extended to an informative relationship with the mathematics of contemporary science.

The oracle Turing machine, which made its first appearance in Turing [42], should be familiar enough. The details are not important, but can be found in most reasonable introductions to computability (see for instance [7]). One just needs to add to the usual picture of a Turing machine the capacity for questioning an oracle set about the membership status of individual natural numbers.

[1]Swift even has a Laputan professor introduce Gulliver to *The Engine*, an (appropriately useless) early anticipation of today's computing machines, and more.

The basic form of the questioning permitted is modelled on that of everyday scientific practice. This is seen most clearly in today's digital data gathering, whereby one is limited to receiving data which can be expressed, and transmitted to others, as information essentially finite in form. But with the model comes the capacity to collate data in such a way as enable us to deal with arbitrarily close approximations to infinitary inputs and hence outputs, giving us an exact counterpart to the computing scientist working with real-world observations. If the different number inputs to the oracle machine result in 0-1 outputs from the corresponding Turing computations, one can collate the outputs to get a binary real computed from the oracle real, the latter now viewed as an input. This gives a partial computable functional Φ, say, from reals to reals, which may sometimes be described as a *Turing reduction*.

As usual, one cannot computably know when the machine for Φ computes on a given natural number input, so Φ may not always give a fully defined real output. So Φ may be partial. One can computably list all oracle machines, and so index the infinite list of all such Φ, but one cannot computably sift out the partial Φ's from the list.

Anyway, put $\mathbb{R}$ together with this list, and we get the Turing Universe. That is, we obtain a structure involving information in the form of real numbers, algorithmically related by all possible Turing reductions. Depending on one's viewpoint, this is either a rather reduced scientific universe (if you are a poet, a philosopher, or a string-theorist), or (if one is vainly looking for the richness of algorithmic content contained on our list in the physical context, being familiar with the richness of emergent structure in the Turing universe) a much expanded one. But we will defer difficult comparisons between the information content of the Turing universe and that of the physical universe until later. For the moment we will follow Emil Post in his search for the informational underpinnings of computational structure in a safer mathematical context.

Post's first step was to gather together binary reals which are computationally indistinguishable from each other, in the sense that they are mutually Turing computable from each other. Mathematically, this delivered a more standard mathematical structure to investigate — the familiar upper semi-lattice of the *degrees of unsolvability*, or *Turing degrees*. There is no simple scientific counterpart of the mathematical model, or any straightforward justification for what Post did with the Turing universe for perfectly good mathematical reasons — if one wants to get a material avatar of the Turing landscape one needs both a closer and a more comprehensive view of the physical context.

5 Definability in Science

Schematically, any causal context framed in terms of everyday computable mathematics can be modelled in terms of Turing reductions. Then emergence can be formalised as definability over the appropriate substructure of the Turing universe; or more generally, as invariance under automorphisms of the Turing universe. Simple and fundamental as the notion of definability is, and basic as it is to everyday thought and discourse, as a concept it is not well understood outside of logic. This is seen most strikingly in the physicists' apparent lack of awareness of the concept in interpreting the collapse of the wave function. Quantum decoherence and the many-worlds hypothesis comprise a far more outlandish interpretive option than does speculating that measurements, in enriching an environment, merely lead to an assertion of definability. It appears a sign of desperation to protect consistent histories by inventing new universes, when the mathematics of our observable universes already contains a straightforward explanation. We have argued (see for instance [13]) that many scientific puzzles can be explained in terms of failures of definability in different contexts, and that the key task is to identify useful theoretical models within which to investigate the nature of definability more fully. One of the most relevant of these models has to be that of Turing, based as it is on a careful analysis of the characteristics of algorithmic computation.

This brings us to a well-known and challenging research programme, initiated by Hartley Rogers in his 1967 paper [36], in which he drew attention to the fundamental problem of characterising the Turing invariant relations. Again, the intuition is that these are key to pinning down how basic laws and entities emerge as mathematical constraints on causal structure. It is important to notice how the richness of Turing structure discovered so far becomes the raw material for a multitude of non-trivially definable relations, matching in its complexity what we attempt to model.

Unfortunately, the current state of Rogers' programme is not good. For a number of years research in this area was dominated by a proposal originating with the Berkeley mathematician Leo Harrington, which can be (very) roughly stated:

Bi-interpretability Conjecture: *The Turing definable relations are exactly those with information content describable in second-order arithmetic.*

Most importantly, bi-interpretability is not consistent with the existence of non-trivial Turing automorphisms. Despite decades of work by a number of leaders in the field, the exact status of the conjecture is still a matter of controversy.

For those of us who have grown up with Thomas Kuhn's 1962 book [29] on the structure of scientific revolutions, such difficulties and disagreements are not seen as primarily professional failures, or triggers to collective shame (although they may be that too), but rather signs that something scientifically important is at stake. A far more public controversy currently shapes developments around important issues affecting theoretical physics — see, for example the recent books of Lee Smolin [40] and Peter Woit [47].

This turns out to be very relevant to our theme of the importance of fundamental notions, such as that of mathematical definability, to the formation of basic scientific theories. In this context, the specific focus on string theory of the above-mentioned books of Smolin and Woit is important, given that string theory was initially intended to remedy a number of inadequacies in current scientific thinking, without really getting to grips with fundamental issues. Our argument is that string theory does very validly point towards a substitution of abstract mathematics for inaccessible observational data. And that it has produced some very beautiful and useful mathematics, and widened our conceptual horizons in relation to models of the universe. But — that it has failed to enlist notions of global definability to pin down important elements of the real world.

As Peter Woit [47, p.1] describes, according to purely pragmatic criteria particle physics has produced a standard model which is remarkably successful, and has great predictive power:

> By 1973, physicists had in place what was to become a fantastically successful theory of fundamental particles and their interactions, a theory that was soon to acquire the name of the standard model. Since that time, the overwhelming triumph of the standard model has been matched by a similarly overwhelming failure to find any way to make further progress on fundamental questions.

The reasons why people are dissatisfied echo misgivings going back to Einstein himself [20, p.63]:

> ... I would like to state a theorem which at present can not be based upon anything more than upon a faith in the simplicity, i.e. intelligibility, of nature ... nature is so constituted that it is possible logically to lay down such strongly determined laws that within these laws only rationally completely determined constants occur (not constants, therefore, whose numerical value could be changed without destroying the theory) ...

If one really does have a satisfying description of how the universe is, it should not contain arbitrary elements with no plausible explanation. In particular, a theory containing arbitrary constants, which one adjusts to fit the intended interpretation of the theory, is not complete. And as Woit observes:

> One way of thinking about what is unsatisfactory about the standard model is that it leaves seventeen non-trivial numbers still to be explained, ...

At one time, it had been hoped that string theory would supply a sufficiently fundamental framework to provide a much more coherent and comprehensive description, in which such arbitrary ingredients were properly pinned down. But despite its mathematical attractions, there are growing misgivings about its claimed status as "the only game in town" as a unifying explanatory theory. Here is how one time string theorist Daniel Friedan [21] combatively puts it:

> The longstanding crisis of string theory is its complete failure to explain or predict any large distance physics. ... String theory is incapable of determining the dimension, geometry, particle spectrum and coupling constants of macroscopic spacetime. ... The reliability of string theory cannot be evaluated, much less established. String theory has no credibility as a candidate theory of physics.

Smolin starts his book [40]:

> From the beginning of physics, there have been those who imagined they would be the last generation to face the unknown. Physics has always seemed to its practitioners to be almost complete. This complacency is shattered only during revolutions, when honest people are forced to admit that they don't know the basics.

He goes on to list what he calls the "five great [unsolved] problems in theoretical physics". Gathering these together, and slightly editing, they are [40, pp.5-16]:

1. Combine general relativity and quantum theory into a single theory that can claim to be the complete theory of nature.
2. Resolve the problems in the foundations of quantum mechanics.
3. The unification of particles and forces problem: Determine whether or not the various particles and forces can be unified in a theory that explains them all as manifestations of a single, fundamental entity.
4. Explain how the values of the free constants in the standard model of physics are chosen in nature.
5. Explain dark matter and dark energy. Or, if they do not exist, determine how and why gravity is modified on large scales.

That each of these questions can be framed in terms of definability is not so surprising, since that is exactly how, essentially, they are approached by researchers. The question is the extent to which progress is impeded by a lack of consciousness of this fact, and an imperfect grip of what is fundamental. Quoting Einstein again (from a letter to Robert Thornton, dated 7 December 1944, Einstein Archive 61-754), this time on the relevance of a philosophical approach to physics:

> So many people today – and even professional scientists – seem to me like someone has seen thousands of trees but has never seen a forest. A knowledge of the historical and philosophical background gives that kind of independence from prejudices of his generation from which most scientists are suffering. This independence created by philosophical insight is – in my opinion – the mark of distinction between a mere artisan or specialist and a real seeker after truth.

Smolin's comment [40, p.263] is in the same direction, though more specifically directed at the string theorists:

> The style of the string theory community ... is a continuation of the culture of elementary-particle theory. This has always been a more brash, aggressive, and competitive atmosphere, in which theorists

> vie to respond quickly to new developments ... and are distrustful of philosophical issues. This style supplanted the more reflective, philosophical style that characterized Einstein and the inventors of quantum theory, and it triumphed as the center of science moved to America and the intellectual focus moved from the exploration of fundamental new theories to their application.

So what is it that is fundamental that is being missed? For Smolin [40, p.241], it is *causality*:

> It is not only the case that the spacetime geometry determines what the causal relations are. This can be turned around: Causal relations can determine determine the spacetime geometry ... Its easy to talk about space or spacetime emerging from something more fundamental, but those who have tried to develop the idea have found it difficult to realize in practice. ... We now believe they failed because they ignored the role that causality plays in spacetime. These days, many of us working on quantum gravity believe that *causality itself is fundamental* – and is thus meaningful even at a level where the notion of space has disappeared.

Citing Penrose as an early champion of the role of causality, he also mentions Rafael Sorkin, Fay Dowker, and Fotini Markopoulou, known in this context for their interesting work on causal sets (see [4]), which abstract from causality relevant aspects of its underlying ordering relation. Essentially, causal sets are partial orderings which are locally finite, providing a model of spacetime with built-in discreteness. Despite the apparent simplicity of the mathematical model, it has had striking success in approximating the known characteristics of spacetime. An early prediction, in tune with observation, concerned the value of Einstein's cosmological constant.

Of course, this preoccupation with causality might suggest to a logician a need to also look at its computational content. Smolin's comment that "Causal relations can determine the spacetime geometry" touches on one of the biggest disappointments with string theory, which turns out to be a 'background dependant' theory with a vengeance — one has literally thousands of candidate Calabi-Yau spaces for shaping the extra dimensions of superstring theory. In current superstring models, Calabi-Yau manifolds are those qualifying as possible space formations for the six hidden spatial dimensions, their undetected status explained by the assumption of their being smaller than currently observable lengths.

Ideally, a truly fundamental mathematical model should be background independent, bringing with it a spacetime geometry arising from within.

6 The Emergence-Definability Symbiosis

There are obvious parallels between the Turing universe and the material world. Each of which in isolation, to those working with specific complexities, may seem superficial and unduly schematic. But the lessons of the history of mathematics and its applications is that the simplest of abstractions can yield unexpectedly far-reaching and deep insights into the nature of the real world. The main achievement of the Turing model, and its definable content, is to illuminate and structure the role of computability theoretic expressions of emergence.

At the most basic level, science describes the world in terms of real numbers. This is not always immediately apparent, any more that the computer on ones desk is obviously an avatar of a universal Turing machine. Nevertheless, scientific theories consist, in their essentials, of postulated relations upon reals. These reals are abstractions, and do not come necessarily with any recognisable metric. They are used because they are the most advanced presentational device we can practically work with. There is no faith that reality itself consists of information presented in terms of reals. In fact, those of us who believe that mathematics is indivisible, no less in its relevance to the material world, have a due humility about the capacity for our science to capture more than a surface description of reality.

Some scientists would take us in the other direction, and claim that the universe is actually finite, or at least countably discrete. We have argued elsewhere (see for example [14]) that to most of us a universe without algorithmic content is inconceivable. And that once one has swallowed that bitter pill, infinitary objects are not just a mathematical convenience (or inconvenience, depending on one's viewpoint), but become part of the mathematical mold on which the world depends for its shape. As it is, we well know how essential algorithmic content is to our understanding of the world. The universe comes with recipes for doing things. It is these recipes which generate the rich information content we observe, and it is reals which are the most capacious receptacles we can humanly carry our information in, and practically unpack.

Globally, there are still many questions concerning the extent to which one can extend the scientific perspective to a comprehensive presentation of the universe in terms of reals — the latter being just what we need to do in order to model the immanent emergence of constants and natural laws from an entire universe. Of course, there are many examples of presentations entailed by scientific models of particular aspects of the real world. But given the fragmentation of science, is fairly clear that less natural presentations may well have an explanatory role, despite their lack of a role in practical computation.

The natural laws we observe are largely based on algorithmic relations between reals. For instance, Newtonian laws of motion will computably predict, under reasonable assumptions, the state of two particles moving under gravity over different moments in time. And the character of the computation involved can be represented as a Turing functional over the reals representing different time-related two-particle states. One can point to physical transitions which are not obviously algorithmic, but these will usually be composite processes, in which the underlying physical principles are understood, but the mathematics of their workings outstrip available analytical techniques. Over forty years ago, Georg Kreisel [27] distinguished between classical systems and *cooperative phenomena* not known to have Turing computable behaviour, and proposed

[28, p.143, Note 2] a collision problem related to the 3-body problem, which might result in "an analog computation of a non-recursive function (by repeating collision experiments sufficiently often)". However, there is a qualitatively different apparent breakdown in computability of natural laws at the quantum level — the *measurement problem* challenges us to explain how certain quantum mechanical probabilities are converted into a well-defined outcome following a measurement. In the absence of a plausible explanation, one is denied a computable prediction. The physical significance of the Turing model depends upon its capacity for explaining what is happening here. If the phenomenon is not composite, it does need to be related in a clear way to a Turing universe designed to model computable causal structure. We will need to talk more about definability and invariance.

For the moment, let us think in terms of what an analysis of the automorphisms of *any* sufficiently comprehensive, sufficiently fundamental, mathematical model of the material universe might deliver.

Let us first look at the relationship between automorphisms and many-worlds. When one says "I tossed a coin and it came down heads, maybe that means there is a parallel universe where I tossed the coin and it came down tails", one is actually predicating a large degree of correspondence between the two parallel universes. The assumption that *you* exist in the two universes puts a huge degree of constraint on the possible differences — but nevertheless, some relatively minor aspect of our universe has been rearranged in the parallel one. There are then different ways of relating this to the mathematical concept of an automorphism. One could say that the two parallel worlds are actually isomorphic, but that the structure was not able to *define* the outcome of the coin toss. So it and its consequences appear differently in the two worlds. Or one could say that what has happened is that the worlds are *not* isomorphic, that actually we were able to change quite a lot, without the parallel universe looking very different, and that it was these fundamental but hidden differences which forces the worlds to be separate and not superimposed, quantum fashion. The second view is more consistent with the view of quantum ambiguity displaying a failure of definability. The suggestion here being that the observed existence of a particle (or cat!) in two different states at the same time merely exhibits an automorphism of our universe under which the classical level is rigid (just as the Turing universe displays rigidity above $\mathbf{0}''$) but under which the sparseness of defining structure at the more basic quantum level enables the automorphism to re-represent our universe, with everything at our level intact, but with the particle in simultaneously different states down at the quantum level. And since our classical world has no need to decohere these different possibilities into parallel universes, we live in a world with the automorphic versions superimposed. But when we make an observation, we establish a link between the undefined state of the particle and the classical level of reality, which destroys the relevance of the automorphism. To believe that we now get parallel universes in which the alternative states are preserved, one now needs to decide how much else one is going to change about our universe to enable the state of the particle destroyed as a possiblity to survive in the parallel universe — and what weird and wonderful things one must accommodate in order to make that feasible. It is hard at this point to discard the benefits brought by a little mathematical sophistication. Quantum ambiguity as a failure of definability is a far more palatable alternative than the invention of new worlds of which we have no evidence or scientific understanding.

Another key conceptual element in the drawing together of a global picture of our universe with a basic mathematical model is the correspondence between emergent phenomena and definable relations. This gives us a framework within which to explain the particular forms of the physical constants and natural laws familiar to us from the standard model science currently provides. It goes some way towards substantiating Penrose's [33, pp106-107] 'strong determinism', according to which "all the complication, variety and apparent randomness that we see all about us, as well as the precise physical laws, are all exact and unambiguous consequences of one single coherent mathematical structure" — and repairs the serious failure of the standard model pointed to by researchers such as Smolin and Woit. It also provides a hierarchical model of the fragmentation of the scientific enterprise. This means that despite the causal connections between say particle physics and the study of living organisms, the corresponding disciplines are based on quite different basic entities and natural laws, and there is no feasible and informative reduction of one to another. The entities in one field may emerge through phase transitions characterised in terms of definable relations in the other, along with their distinct causal structures. In this context, it may be that the answer to Smolin's first 'great unsolved problem in theoretical physics' consists of an explanation of why there is no single theory (of the kind that makes useful predictions) combining general relativity and quantum theory.

For further discussion of such issues, see [6], [9], [10], [11], [13] and [14].

References

[1] I. Adler, D. Barab and R. V. Jean, A history of the study of phyllotaxis, *Annals of Botany*, **80** (1997), 231–244.

[2] R. C. Arkin, *Behaviour-Based Robotics,* M.I.T. Press, Cambridge, MA, 1998.

[3] E. J. Beggs and J. V. Tucker, Experimental computation of real numbers by Newtonian machines, to appear in *Proceedings of the Royal Society Series A*.

[4] L. Bombelli, J. Lee, D. Meyer and R. D. Sorkin, Spacetime as a causal set, Phys. Rev. Lett. **59** (1987), 521–524.

[5] R. Brooks, The relationship between matter and life. *Nature* **409** (2001), 409–411.

[6] S. B. Cooper, Clockwork or Turing U/universe? - Remarks on causal determinism and computability, in *Models and Computability* (S. B. Cooper and J. K. Truss, eds.), London Mathematical Society Lecture Notes Series 259, Cambridge University Press, Cambridge, New York, Melbourne, 1999, pp.63–116.

[7] S. B. Cooper, *Computability Theory,* Chapman & Hall/CRC, Boca Raton, London, New York, Washington, D.C., 2004.

[8] S. B. Cooper, Computability and emergence. In *Mathematical Problems from Applied Logic. New Logics for the XXI-st Century II* (eds. Gabbay D., Goncharov S., Zakharyaschev M.) Kluwer/Springer International Mathematical Series, Vol. 5, 2005.

[9] S. B. Cooper, Definability as hypercomputational effect, Applied Mathematics and Computation, **178** (2006), 72–82.

[10] S. B. Cooper, How Can Nature Help Us Compute?, in *SOFSEM 2006: Theory and Practice of Computer Science - 32nd Conference on Current Trends in Theory and Practice of Computer Science, Merin, Czech Republic, January 2006* (J. Wiedermann, J. Stuller, G. Tel, J. Pokorny, M. Bielikova, eds.), Springer Lecture Notes in Computer Science No. 3831, 2006, pp.1–13.

[11] S. B. Cooper, Computability and emergence, in *Mathematical Problems from Applied Logic I. Logics for the XXIst Century* (D.M. Gabbay, S.S. Goncharov, M. Zakharyaschev, eds.), Springer International Mathematical Series, Vol. 4, 2006, pp. 193–231.

[12] S. B. Cooper, The Incomputable Alan Turing, in the *Proceedings of Alan Mathison Turing 2004: A celebration of his life and achievements, Manchester University, 5 June, 2004* (Janet Delve and Jeff Paris, eds.), electronically published by the British Computer Society (2008): http://www.bcs.org/server.php?show=nav.9917.

[13] S. B. Cooper, Extending and interpreting Post's Programme, to appear.

[14] S. B. Cooper and P. Odifreddi, Incomputability in Nature, in *Computability and Models* (S.B. Cooper and S.S. Goncharov, eds.), Kluwer Academic/Plenum, New York, Boston, Dordrecht, London, Moscow, 2003, pages 137–160.

[15] J. Copeland, Turing's O-machines, Penrose, Searle, and the brain. *Analysis*, **58** (1998), 128–38.

[16] J. Copeland and D. Proudfoot, On Alan Turing's anticipation of connectionism. *Synthese*, **108** (1996), 361–377. Reprinted in *Artificial Intelligence: Critical Concepts in Cognitive Science* (R. Chrisley, ed), Volume 2: Symbolic AI. Routledge, London, 2000.

[17] A. Damasio, *The Feeling Of What Happens.* Harcourt, Orlando, FL., 1999.

[18] M. Davis (Ed.), *Solvability, Provability, Definability: The Collected Works of Emil L. Post*, Birkhäuser, Boston, Basel, Berlin, 1994.

[19] Deutsch D. (1997) *The Fabric of Reality.* Penguin, London, New York.

[20] A. Einstein, Autobiographical Notes, in *Albert Einstein: Philosopher-Scientist* (P. Schilpp, ed.), Open Court Publishing, 1969.

[21] D. Friedan, A Tentative Theory of Large Distance Physics, J. High Energy Phys. JHEP10 (2003) 063.

[22] R. Geroch and J. B. Hartle, Computability and physical theories, *Foundations of Physics* **16** (1986), 533–550.

[23] D. Goldin and P. Wegner, Computation Beyond Turing Machines: seeking appropriate methods to model computing and human thought. *Communications of the ACM* **46** (2003), 100–102.

[24] D. Goldin and P. Wegner, The Church-Turing Thesis: Breaking the Myth, in *CiE 2005: New Computational Paradigms: Papers presented at the conference in Amsterdam, June 8–12, 2005* (Cooper S. B., Löwe B., Torenvliet L., eds.), Lecture Notes in Computer Science No. 3526, Springer-Verlag, 2005.

[25] A. Hodges, *Alan Turing: The Enigma,* Vintage, London, Melbourne, Johannesburg, 1992.

[26] S. A. Kauffman, *Reinventing the Sacred: A New View of Science, Reason and Religion,* Basic Books, New York, 2008.

[27] G. Kreisel, Mathematical logic: What has it done for the philosophy of mathematics?, in *Bertrand Russell, Philosopher of the Century* (R. Schoenman, ed.), Allen and Unwin, London, 1967, pp. 201–272.

[28] G. Kreisel, Church's Thesis: a kind of reducibility axiom for constructive mathematics, in *Intuitionism and proof theory: Proceedings of the Summer Conference at Buffalo N.Y. 1968* (A. Kino, J. Myhill and R.E. Vesley, eds.), North-Holland, Amsterdam, London, 1970, pp. 121–150.

[29] T. S. Kuhn, *The Structure of Scientific Revolutions*, Third edition 1996, University of Chicago Press, Chicago, London.

[30] R. Milner, Elements of interaction:Turing award lecture. *Communications of the ACM* **36** (1993), 78–89.

[31] D. Knuth, *The Art of Computer Programming, Vol. 1: Fundamental Algorithms*, Addison-Wesley, 1968.

[32] W. McCulloch and W. Pitts, A logical calculus of the ideas immanent in nervous activity, *Bull. Math. Biophys.* **5** (2003), 115–133.

[33] R. Penrose, Quantum physics and conscious thought, in *Quantum Implications: Essays in honour of David Bohm* (B.J. Hiley and F.D. Peat, eds.), Routledge & Kegan Paul, London, New York, 1987, pp. 105–120.

[34] S. Pinker, *How the Mind Works.* W. W. Norton, New York, 1997.

[35] E. L. Post, Recursively enumerable sets of positive integers and their decision problems, *Bulletin of the American Mathematical Society*, **50** (1944), 284–316; reprinted in E. L. Post, *Solvability, Provability, Definability: The Collected Works of Emil L. Post*, pages 461–494.

[36] H. Rogers, Jr., Some problems of definability in recursive function theory, in *Sets, Models and Recursion Theory* (J.N.Crossley, ed.), Proceedings of the Summer School in Mathematical Logic and Tenth Logic Colloquium, Leicester, August–September, 1965, North-Holland, Amsterdam, pp. 183–201.

[37] E. M. A. Ronald, M. Sipper and M. S. Capcarrère, Design, observation, surprise! A test of emergence, Artificial Life, **5** (1999), 225–239.

[38] D. Ruelle and F. Takens, On the nature of turbulence, *Communications of Mathematical Physics*, **20** (1971), 167–192.

[39] P. Smolensky, On the proper treatment of connectionism, *Behavioral and Brain Sciences* **11** (1988), 1–74.

[40] L. Smolin, *The Trouble With Physics: The Rise of String Theory, the Fall of Science and What Comes Next,* Allen Lane/Houghton Mifflin, London, New York, 2006.

[41] A. Turing, On computable numbers, with an application to the Entscheidungsproblem, *Proceedings of the London Mathematical Society*, Vol. 42, 1936–37, pages 230–265; reprinted in A. M. Turing, *Collected Works: Mathematical Logic*, pages 18–53.

[42] A. Turing, Systems of logic based on ordinals, *Proceedings of the London Mathematical Society*, Vol. 45, 1939, pages 161–228; reprinted in A.M. Turing, *Collected Works: Mathematical Logic*, pages 81–148.

[43] A. M. Turing, *Intelligent machinery.* National Physical Laboratory Report, 1948. In *Machine Intelligence 5* (Meltzer B. and Michie D., eds.), Edinburgh University Press, Edinburgh, 1969, pp. 3–23. Reprinted in A. M. Turing, *Collected Works: Mechanical Intelligence* (Ince D. C., ed.), North-Holland, Amsterdam, New York, Oxford, Tokyo, 1992.

[44] A. M. Turing, The chemical basis of morphogenesis, *Philosophical Transactions of the Royal Society of London*, **B 237** (1952), 37–72; reprinted in A.M. Turing, *Collected Works: Morphogenesis*.

[45] A. M. Turing, *Collected Works, Vol. 3: Morphogenesis* (P. T. Saunders, Editor), Elsevier, Amsterdam, New York, Oxford, Tokyo, 1992.

[46] A. M. Turing, *Collected Works: Mathematical Logic* (R. O. Gandy and C. E. M. Yates, Editors), Elsevier, Amsterdam, New York, Oxford, Tokyo, 2001.

[47] P. Woit, *Not Even Wrong: The Failure of String Theory and the Continuing Challenge to Unify the Laws of Physics,* Jonathan Cape, London, 2006.