

FROM DESCARTES TO TURING: THE COMPUTATIONAL CONTENT OF SUPERVENIENCE

S. Barry Cooper*
University of Leeds, Leeds LS2 9JT, U.K.
pmt6sbc@leeds.ac.uk

Abstract

Mathematics can provide precise formulations of relatively vague concepts and problems from the real world, and bring out underlying structure common to diverse scientific areas. Sometimes very natural mathematical concepts lie neglected and not widely understood for many years, before their fundamental relevance is recognised and their explanatory power is fully exploited. The notion of definability in a structure is such a concept, and Turing's [77] 1939 model of interactive computation provides a fruitful context in which to exercise the usefulness of definability as a powerful and widely applicable source of understanding. In this article we set out to relate this simple idea to one of the oldest and apparently least scientifically approachable of problems — that of realistically modelling how mental properties supervene on physical ones.

Mathematics can provide precise formulations of relatively vague concepts and problems from the real world, and bring out underlying structure common to diverse scientific areas. Sometimes very natural mathematical concepts lie neglected and not widely understood for many years, before their fundamental relevance is recognised and their explanatory power is fully exploited. Previously we have argued that the notion of definability in a structure is such a concept, and pointed to Turing's [77] 1939 model of interactive computation as providing a fruitful context in which to exercise the usefulness of definability as a powerful and widely applicable source of understanding.

Below, we relate this simple idea to one of the oldest and apparently least scientifically approachable of problems — that of realistically modelling how mental properties supervene on physical ones. We will first briefly review the origins with René Descartes of mind-body dualism, and the problem of mental causation. We will then summarise the subsequent difficulties encountered, and their current persistence, and the more recent usefulness of the concept of *supervenience* in

*Research supported by EPSRC research grant No. *Computing with Partial Information: Definability in the Local Structure of the Enumeration Degrees*.

providing a philosophical workspace in which to make mind-body connections — and the parallel recognition of *emergence* as a tool for giving supervenience a non-reductive physical content. The rise and fall of the British emergentists will provide a salutary warning of the pitfalls of working with too vague a formulation of emergence, and of the need for a test for emergence.

Following on from this, we will further clarify emergence by looking at mathematical analogues, and at approaches to emergence — such as synergetics — with a strong mathematical or scientific content. This will move us away from the empirical quest for emergent phenomena as something surprising towards formalisations of self-organisation in terms of the two-way interaction between physical phenomena, and their descriptions and representations. We will point to the important role of representation in building computationally complex systems with the capability to transcend the Turing barrier. Finally, we will argue for the fundamental role of causality and that of the extended Turing model of the algorithmic content of nature. What we know about the theory of this model will then lead to a clarifying return to the problems of fragmentation and coherence first highlighted by Descartes. In doing so, we will point to the existence of different levels of ‘causality’, and to give mathematical substance to the common distrust of causality as a term in everyday usage.

1 SUBSTANCE DUALISM, AND THE PROBLEM OF MENTAL CAUSATION

In general, Descartes aimed at explanations in terms of strictly mechanical principles and mathematical models. But living with an experience of mental phenomena for which there is no obvious underlying mechanism, he proclaimed the essential non-physicality of the mind — for instance, in the Discourse on Method, Part IV, from 1637:

I [am] a substance the whole essence or nature of which is to think, and for its existence there is no need of any place, nor does it depend on any material thing; so that this me, that is to say, the soul by which I am what I am, is entirely distinct from body . . .

Whereas, regarding the body itself (Discourse on Method, Part V), he says:

. . . this will not seem strange to those, who, knowing how many different *automata* or moving machines can be made by the industry of man, without employing in so doing more than a very few parts in comparison with the great multitude of bones, muscles, nerves, arteries, veins, or other parts that are found in the body of each animal. From this aspect the body is regarded as a machine which, having been made by the hands of God, is incomparably better arranged, and possesses in itself movements which are much more admirable, than any of those which can be invented by man.

So Descartes views the body as a very complicated and well-designed machine, and thought and free will as being quite separated from it in nature. As is well-known, Descartes' 'substance dualism' was not well-received, even at the time. Though notwithstanding this, the historical and current failure to build man-made machines to even rival animals (whom Descartes thought certainly would be mechanical in nature), suggests that there is *something* in need of further explanation — and not just in terms of the cleverness of God at constructing automata.

There are various problems connected with such a dualist reading of the relationship between mind and body. The chief of these is the problem of mental causation: How can mentality have a causal role in a world that is fundamentally physical? And then there is the related problem of 'overdetermination' — the problem of phenomena having, according to how one views them, both mental and physical causes. According to Jaegwon Kim [35, p.156]: "What has become increasingly evident over the past thirty years is that mental causation poses insuperable difficulties for all forms of mind-body dualism ...". Unless of course one takes the epiphenomenalist view, its origins associated with T.H. Huxley [32], and partly supported by neuroscience research over the years, which holds that mental phenomena, though physically caused, do not themselves have a causal impact on the physical world. Otherwise, these problems outlined by Kim persist, in different guises, even within all but the most openly reductive attempts at materialist interpretations. Here is Kim [35, p.1] again:

... the problem of mental causation is solvable only if mentality is physically reducible; however, phenomenal consciousness resists physical reduction, putting its causal efficacy in peril.

And more explicitly [35, p.1]:

How can the mind exercise its causal powers in a causally closed physical world? Why is there, and how can there be, such a thing as the mind, or consciousness, in a physical world? ... these two problems, mental causation and consciousness, are intertwined, and that, in a sense, they make each other insoluble.

In fact, Descartes opened up a Pandora's (black) box of more-or-less weird and wonderful takes on the question of how to characterise the mind-body relationship in a scientifically constrained world. How do we characterise the link between mind and physical world? And achieve this in a way that respects the complexities involved?

The spectrum of subtle differences in approach is largely contained within the reductive–nonreductive physicalism spectrum. But the more one tries to clarify the notions involved, the more difficulties one encounters. One is reminded of the old Indian tale of a group of blind men who separately touch different parts of an elephant to learn what it is like, each coming up with very different conclusions based on their particular relationship to the elephant, and on what they were previously familiar with. The version captured in John Godfrey Saxe's 19th century poem *The Blind Men and the Elephant* concludes:

MORAL.

So oft in theologic wars,
The disputants, I ween,
Rail on in utter ignorance
Of what each other mean,
And prate about an Elephant
Not one of them has seen!

Donald Davidson's anomalous monism (see, for example, [17]) establishes just one outpost of this unruly debate. While this is not the right place to discuss this in detail, it is worth remarking that it exploits the tension between events being causally connected, and their being explainable in terms of well-understood laws governing the causal relationship. This distinction is familiar to the computationally-minded, and re-emerges below when we try to extract the computational content of such elusive processes as those governing the mind.

Towards the other end of the spectrum, one finds Saxe's 'MORAL' used to downplay the existence of such entities as consciousness and qualia, which continue to bother observers such as Kim. Here is Daniel Dennett [21, pp.369-370]:

It's not hard to see how philosophers have tied themselves into such knots over qualia. They started where anyone with any sense would start: with their strongest and clearest intuitions about their own minds. Those intuitions, alas, form a mutually self-supporting closed circle of doctrines, imprisoning their imaginations in the Cartesian Theater. Even though philosophers have discovered the paradoxes inherent in this closed circle of ideas — that's why the literature on qualia exists — they haven't had a *whole alternative vision* to leap to, and so, trusting their still-strong intuitions, they get dragged back into the paradoxical prison.

But then, as John Searle [65, p.102] observes, "If Dennett denies the existence of conscious states as we usually think of them, what is his alternative account? Not surprisingly, it is a version of Strong AI." In this way, the relevance of suitably modelling the computational content underlying such controversies is strikingly brought home to us.

2 SUPERVENIENCE AS WORKSPACE

Dennett's view of consciousness is often described as being a *functionalist* one. In the context of the philosophy of mind, a functionalist will look for explanations of mental states in terms of the underlying causal relationships of the system, rather than in the particular physics or biology of its constitution. It is an approach to mental properties which underpins most of cognitive science, and is specially associated with the logician Hilary Putnam (see [55], [56]). Particularly influential

was Turing's 1950 paper [79], replacing the question "Can machines think?" with that of whether it is possible for a machine to pass the Turing Test.

More recently, within computer science the virtual machine paradigm, going back to the 1974 paper [51] of Popek and Goldberg, is seen to be very relevant. As commented by Aaron Sloman (private communication):

...the relations between (a) running virtual machines in computers and (b) the processes in underlying physical machines, ... at one level ... are well understood as they are created, debugged, extended, and documented by human engineers, yet they have many of the features (including causal over-determination) that have caused puzzlement when noticed in much more complex systems produced by biological evolution (e.g. human minds and brains) and others (e.g. socio-economic systems implemented in psychological systems or the underlying physical processes: a credit crunch seems to be a process in a socio-economic virtual machine distributed over the surface of the earth.)

Sloman has written interestingly on the relations between running virtual machines and the physical implementations, recently taking issue with the Brooks' well-known paper [10] in [68].

What is attractive about functionalism is the freedom it gives to us to seek aspects of mentality beyond the human context, for example in animals or machines. On the other hand, as an explanation of qualia or our experience of consciousness its abstractions seem inadequate to many, based on failures of modelling of the structural complexities involved — or, to put it more simply, a lack of respect for the real world. Functionalism is certainly an attractive perspective for the computer scientist, appearing to free the computational process from a dependence on the medium hosting it. The functionalist may validly draw useful support for a computational perspective on mental phenomena. But in its assumption of the primacy of the functional relationship, it makes it harder to model the infrastructure of such relationships, an understanding of which the character of the physical remains the key. The potential consistency of functionalism with different ends of the materialism-dualism spectrum seems to signal that something is missing from its explanatory power. This how William Hasker expresses it [28, p.32]:

... either the causal-functional states identified by functionalism characteristically involve conscious "feels" (in the case of qualia) and "aboutness" (in the case of intentional states) or they do not. If they do not, then functionalism is not a theory about the mental at all; the name remains, but the subject has been changed. If they do, then functionalism may indeed be a theory (possibly even a true theory) about the mental, but it does nothing to explain two of the most salient features of the mental states in question, namely subjective consciousness and intentional reference. In this case, there seems to be a need for a theory that more directly addresses the key issue concerning the relationship between consciousness and its physical embodiment.

The problem is that functionalism, much as it *fits* with what follows, does not *provide a workspace* for addressing its own explanatory deficiencies. And this is where the notion of supervenience, as understood in current philosophical usage, comes in. According to Kim [34, 14-15], supervenience:

...represents the idea that mentality is at bottom physically based, and that there is no free-floating mentality unanchored in the physical nature of objects and events in which it is manifested.

There are different ways of formalising the notion, but we can say that:

A set of properties A supervenes upon another set B just in case no two things can differ with respect to A-properties without also differing with respect to their B-properties. (Stanford Encyclopedia of Philosophy)

Then the question becomes: If mental properties supervene on physical properties – how do the complexities of the physical world match those of the mental – and, potentially, of the mirroring process? To quote Aaron Sloman (private communication) again:

Unfortunately, most philosophers discussing supervenience of mind on matter don't pay detailed attention to the variety of coexisting, enduring, interacting structures and processes that exist in a mind (as in virtual machines in computers), so they ignore most of the complexity in the problem of supervenience.

The rest of this article goes some way to support this viewpoint.

3 THE ROOTS OF REDUCTIVE PHYSICALISM

Newton himself would probably have been surprised at the way he changed the way people, and particularly scientists, viewed the world. In *Isaac Newton – The Last Sorcerer* [82], Michael White describes the key role alchemy and religion played in Newton's thinking, and Newton's obsession with biblical prophecy as found in his *Observations upon the Prophecies of Daniel*. He comments:

This demonstrates . . . the radical difference between the intellectual perspectives of the seventeenth and twentieth centuries . . .

Newton set new standards of descriptive control over natural phenomena, which were to fuel great scientific advances, as well as a distrust of the power and validity of less precise ways of thinking. From Galileo and Newton onwards, the overarching aim of science became the extraction of the mechanical content of the world — a 'clockwork universe'. This development was very much in tune with the various shades of materialism espoused earlier by Descartes, Bacon and Hobbes (most notably in *The Leviathan* from 1651), amongst others. But it was Newton who substantiated the mathematical drive to describe the world in terms of well-defined

relations over the continuum, in such a way that predictions emerged as computable functions or simulations, so powerfully dispelling a large range of natural mysteries.

On the other hand, it is well-known that the algorithmic content of a mathematical description involving quite a small number of quantifiers can be largely lost to us. A faith in the descriptive counterpart of nature being logically simple enough to be universally subject to straightforward empirical test is not very scientific, but then the logic that tells us this is less than a hundred years old. When theories such as general relativity and quantum field theory present empirical and mathematical problems of a new kind, one needs to face the fact that our scientific model is in need of an upgrade.

This model, built on the striking success of Newton in predicting such events as the motions of the planets, we recognise in Laplace's description (in [39]) of his 'predictive demon':

Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situations of the beings who compose it – an intelligence sufficiently vast to submit these data to analysis – it would embrace in the same formula the movements of the greatest bodies and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes.

There was very little precision to this model, and there would not be for more than a hundred years after the death of Pierre-Simon, Marquis de Laplace. But there is little doubt that Laplace's demon became in a very real sense an aspirational model for generations of scientists, and one which still holds a strong grip over the thinking of not just scientists, but people from many different walks of life. One can even detect it in the wider social context, when one finds [24] Engels writing in 1880:

...modern materialism sees in [history] the process of evolution of humanity, and aims at discovering the laws thereof. ... [it] embraces the more recent discoveries of natural science, according to which Nature also has its history in time, the celestial bodies, like the organic species that, under favorable conditions, people them, being born and perishing. [with] two great discoveries, the materialistic conception of history and the revelation of the secret of capitalistic production through surplus-value ... Socialism became a science. The next thing was to work out all its details and relations.

So the aim was to work out the "details and relations", the assumption being that the result would provide a possibly complicated, but mechanistic, model of the real world. And that the predictive power contained therein would allow one to not just understand the course of history, but to manipulate and predict it — the machine would be programmable. This would allow no room for an interactive Cartesian dualism. Echoing Gilbert Ryle's dismissal of any such thing — "I shall

often speak of it, with deliberate abusiveness, as the dogma of the Ghost in the Machine” [63, ch.1] — the machine would be effectively ‘ghost-free’.

The first appearance of a mathematical model of such a machine was in two independent and almost simultaneous papers, by Emil Post [52] and Alan Turing [76], describing equivalent, and remarkably similar, formulations of the standard Turing machine model of computation. Such machines only deal with discrete data, so need a little extra work for them to model computation over continuous data, or in a scientific context. Nevertheless, they play an important conceptual role in buttressing the Laplacian model of a potentially deterministic, and predictably so, real world. And, given that the essentials of many computational contexts can be captured perfectly well with approximated data, they provide a model of computable natural processes with a wide range of valid application.

However, any structure with algorithmic content entails some infinitary mathematics, which may involve new relations with physical significance. Turing’s Universal Machine first seen in his 1936 paper [76] not only anticipated today’s stored program computer, even feeding into von Neumann’s thinking on computer design — there was also a very significant ‘ghost in the machine’, that of *incomputability*. Turing’s coding up of computing machines, in such a way that they could be used as input-data, also enabled a simple and quite effective diagonalisation leading to an incomputable real number.

4 NATURE AS DISCIPLINE PROBLEM

There was no real evidence though that incomputability was any more than a mathematical curiosity — and for many people, there still isn’t. How does one distinguish a computational task which is very complicated, to the point where it outstrips the capabilities of our most powerful computers and mathematical resources, from one for which there is an absolute theoretical barrier to it being computationally captured?

Philosophically this is an important question, since its answer will bring with it mathematical baggage; and how we characterize the *material* nature of the world will need to reflect and accommodate the underlying *mathematical* constraints. In practice, though, one can still argue that the *predictive* power of the mathematics is zero, so that despite the philosophy, it remains a curiosity. This view could only be challenged by a development of the theory of incomputability to a point where it can fill a need for a new explanatory framework; a framework with predictive power. And even then, the Laplacian model brings with it a strong syzygial relationship between machine and material which is hard to break. In fact, the survival of the model depends on this. And moving from the security of the Laplacian model requires a whole Kuhnian shift of paradigm, made harder when the new paradigm appears difficult to understand. At first sight, this is not strictly relevant to the problem of characterising the link between mind and brain; that is, until one recognises that the mathematics of different scenarios necessarily brings with it differing material avatars with correspondingly different functionalities an

properties — at some level.

There are many examples of non-trivial reductions of phenomena to the Turing model. This works best when one has an agreed model for the phenomenon in question, which enables something close to a mathematical proof which one can share with others. A well-known example is the 1985 [23] reduction by David Deutsch of the standard model of quantum computation to the Turing model, while proposing the first universal quantum Turing machine. But for Deutsch, this reduction seems to be part of a wider reductive programme, when we see in *Question and Answers with David Deutsch*, on the New.Scientist.com News Service, December, 2006:

I am sure we will have [conscious computers], I expect they will be purely classical, and I expect that it will be a long time in the future. Significant advances in our philosophical understanding of what consciousness is, will be needed.

Another well-known reductionist – paradoxically, most famous for his part in showing how simple high-school mathematics leads to a diversity of incomputability – is Martin Davis. Here is another far-reaching claim, taken from his article [20]:

The great success of modern computers as all-purpose algorithm-executing engines embodying Turing’s universal computer in physical form, makes it extremely plausible that the abstract theory of computability gives the correct answer to the question What is a computation?, and, by itself, makes the existence of any more general form of computation extremely doubtful.

But back in the real world, one sees persistent problems of prediction, and even of description: how does one get beyond the probabilities governing quantum phenomena associated with measurements; how does one capture emergent phenomena in a general predictive framework; and related to this, how to improve our understanding of chaotic/turbulent contexts and their ‘strange attractors’; or to reduce relativistic systems such as black holes to suitably Turing-like models. Even at the level of practical computation, there is a growing interest in non-standard computational models, including a renewed interest in analog and hybrid computing machines – presumably arising from a view, put in writing by J. van Leeuwen and J. Wiedermann [81] that:

... the classical Turing paradigm may no longer be fully appropriate to capture all features of present-day computing.

How does one mathematically deconstruct such apparently complex physical examples? A pointer to the underlying simplicity of the mathematics involved is found in comments of Georg Kreisel [37] from 1970, where he proposes a collision problem related to the 3-body problem, which might result in “an analog computation of a non-recursive function”. What is important in Kreisel’s example is his notion of *co-operative phenomenon*, where the apparent simplicity masks a

nonlinearity and an unavoidably infinitary mathematical analysis, a prerequisite for the mathematics of incomputability to play a role. A more recent example, in which the infinitistic nature of physical phenomena emerges much more explicitly, is provided by the solution to the Painlevé Problem from 1895, asking whether noncollision singularities exist for the N-body problem for $N \geq 4$. For $N = 5$, Jeff Xia showed in 1988 that the answer is Yes (see Saari and Xia [64]). Other work in a classical Newtonian setting includes recent papers by Tucker and Beggs and their collaborators (see, for example, [3], [4]).

Of course, Kreisel thought deeply about extensions of the Church-Turing thesis, and one can find a more than twenty-page discussion related to his thinking on this [46, pp.101–123] in volume 1 of Piergiorgio Odifreddi’s book on Classical Recursion Theory. A valuable and more detailed account of Kreisel’s views can be found in Odifreddi’s article [47] on *Kreisel’s Church*.

5 CHAOS AND ITS ANALOGUES

In situations where we find it hard to mathematically capture natural phenomena, we can sometimes learn a lot from reasonably well-understood mathematical analogues of them. Chaos theory typically deals with situations where the generation of informational complexity is via quite simple rules, but is accompanied by the emergence of new and surprising regularities. A classic example is that of Robert Shaw’s dripping taps (see [67],[66]). What could be dynamically simpler, we think? But the emergent ‘strange attractors’ are interesting and have a less-than-simple relationship to the underlying basic laws, certainly not one that is known to be practically computable.

There is a familiar metaphorical link between complexity arising from basic laws in nature, and between mathematical objects, such as the Mandelbrot and Julia sets. The mathematics attached to such well-defined structures may not be directly applicable to the rather messier physical situations, but the parallel between the basic elements on each side — the simple basic laws, the complexity arising from them, the surprising *emergence* (a key notion) of new formal regularities — is close enough for a mathematical overview of the ‘toy complexity’ to add important new elements to our understanding of what is happening in nature. In particular, there are basic questions concerning the computability or otherwise of the Mandelbrot and Julia sets, which have attracted the attention of high-profile figures, such as Stephen Smale and Roger Penrose, who arrive with impressive track-records as researchers in areas not directly related to computability theoretic questions. This is how Penrose sets the scene in *The Emperors New mind* [49] for the basic question about computability:

Now we witnessed . . . a certain extraordinarily complicated looking set, namely the Mandelbrot set. Although the rules which provide its definition are surprisingly simple, the set itself exhibits an endless variety of highly elaborate structures.

Blum and Smale [6] have actually used the BSS model of computation over arbi-

trary commutative rings (described in [5]) to show the Mandelbrot set to be incomputable; but as Brattka [7] points out, the BSS model gives the incomputability of such intuitively computable sets as the epigraph of the exponential function.

Of course, one gets the Mandelbrot set as the set of all complex numbers c for which the corresponding transform $z \rightarrow z^2 + c$ stays finite, despite unbounded iterations starting with $z = 0$. So at first sight the recognition of a c in the set involves a two-quantifier expression. But on closer inspection, one finds that the complement of the Mandelbrot set can be expressed using just an existential quantifier (a fact that the generation of computer-screen images of the set depends on), so that this complement is intuitively ‘computably enumerable’. Put in this light, it is not so surprising that we are looking at possible incomputability.

What we learn from this is to take the logical structure of scientific descriptions a little more seriously, and to look for implications for the computational — and hence predictive — content of particular theories and their applications in the real world. These may be crude projections we are constructing; but arguably less crude than those widely used to prop up the Laplacian model of science, and related extensions of the Church-Turing thesis. And we will see that they contain enough explanatory power to add to the impression that we have what Thomas Kuhn describes as a ‘period of tension’ characteristic of a prospective paradigm change.

Another important observation is that there seems to be an association between the logical complexity of the description of the Mandelbrot set and the incomputability it promises. And a further observable symptom of these mathematically abstract characteristics, in the form of *emergent* visually beautiful and endlessly unfolding inner structure. This is why such examples are more useful than, say, those provided by the negative solution to Hilbert’s Tenth Problem, and its delivery of diophantine expressions for all computably enumerable sets. It is the memory of the very visual emergence exhibited by the Mandelbrot and Julia sets that gives us pause for thought as we watch the endless variety of wave formations breaking on the shore, or the patterns of clouds morphing across the sky. We now have a very detailed understanding of the scientific laws acting locally within, say, the Cat’s Eye Nebula, but we look with some wonder and a sense of humility at the intricate global structure observed via sufficiently large and well-placed telescopes. Emergence occurs everywhere, and it presents important conceptual and mathematical challenges.

These challenges take on a new dimension in the next section, when one attempts to apply our lessons the task in hand, that of clarifying the link between mental properties and their physical host, the brain. But for Daniel Hillis (Chief Technology Officer of Applied Minds, Inc., and ex-Vice President, Research and Development at Walt Disney Imagineering) even the basic and very down-to-earth task of building intelligent machines involves an intimate relationship with emergence (from *Red Herring Magazine*, April 2001):

I used to think we’d do it by engineering. Now I believe we’ll evolve them. We’re likely to make thinking machines before we understand

how the mind works, which is kind of backwards.

6 RISE AND FALL OF BRITISH EMERGENTISM

In recent years, emergence has figured increasingly in attempts to develop a non-reductive physicalist explanation of our experience of higher mental functionality. But there were earlier anticipations of this contemporary trend, particularly those associated with people working around the end of the nineteenth and beginning of the twentieth century, who became grouped under the heading of ‘British emergentism’. The earliest and best-known of these was John Stuart Mill. Although the term ‘emergence’ was yet to achieve its later currency, the developing sense of its meaning and scope is evident in this quote from John Stuart Mill (*A System of Logic*, Bk.III, Ch.6, 1):

All organised bodies are composed of parts, similar to those composing inorganic nature, and which have even themselves existed in an inorganic state; but the phenomena of life, which result from the juxtaposition of those parts in a certain manner, bear no analogy to any of the effects which would be produced by the action of the component substances considered as mere physical agents. To whatever degree we might imagine our knowledge of the properties of the several ingredients of a living body to be extended and perfected, it is certain that no mere summing up of the separate actions of those elements will ever amount to the action of the living body itself.

For a full account of the history of the British emergentists one should consult Brian McLaughlin’s fascinating survey [44] *The rise and fall of British emergentism*. In the context of this article, the history becomes not just a pointer to the robustness of the intuitions involved, but – more importantly – a cautionary tale centred on an awareness of the reasons for the early emergentists’ decline.

In regard to the former, one finds Samuel Alexander reaching for the sort of explanatory power we see in more recent work, when he talks [1, p. 45] about “the emergence of the quality of consciousness from a lower level of complexity”, or again [1, p. 14]:

The argument is that mind has certain specific characters to which there is or even can be no neural counterpart . . . Mind is, according to our interpretation of the facts, an ‘emergent’ from life, and life an emergent from a lower physico-chemical level of existence.

One of the more modern figures in the group active around the 1920s was C.D. Broad, his views finding full expression in his 1923 book [8] on *The Mind and Its Place In Nature*. Here is his take on what emergence has to contribute to the physicalist agenda [8, p.623]:

. . . the mental properties of those events which do have mental properties are completely determined by the material properties which these

events also have ... it is certainly not ... a form of *Reductive* Materialism; it is a form of the theory ... of *Emergent* Materialism.

Within their differing approaches, the emergentists of the 1920s, such as Broad, Samuel Alexander, and C. Lloyd Morgan, conceived of emergence in ways we would recognise today. Schematically, they had identified some of the distinctive characteristics of an emergent property or phenomenon, and had a recognisably modern check list of explanatory tasks they wanted emergence to perform. In particular, Broad and others saw it as being relevant to a number of reductive failures in science, from the mind-body problem to the apparent irreducibility of the 'special sciences' to more basic sciences, and the nature of the relationship of biology to chemistry, and of these to physics. Emergence supported a familiar hierarchical, 'layered' view of nature and its science.

But there was no precise *test* for emergence, neither via practical observation, nor via analysis of a mathematically valid modelling. Emergence was a useful term for those who shared a common appreciation of the existence of what it referred to, but, within the then state of scientific theory, could be misapplied. Here is Broad explaining what emergence is, and trying to convince the reader that it really does exist, and is something non-reductive and different [8, p.59]:

... the characteristic behaviour of the whole [system] *could* not, even in theory, be deduced from the most complete knowledge of the behaviour of its components, taken separately or in other combinations, and of their proportions and arrangements in this whole. This ... is what I understand by the "Theory of Emergence". I cannot give a conclusive example of it, since it is a matter of controversy whether it actually applies to anything ... I will merely remark that, so far as I know at present, the characteristic behaviour of Common Salt cannot be deduced from the most complete knowledge of the properties of Sodium in isolation; or of Chlorine in isolation; or of other compounds of Sodium, such as Sodium Sulphate ...

So, as we start reading the above quotation, we are nodding our heads in agreement; we are then mildly interested to see Broad's acknowledgement that in his time it was 'a matter of controversy whether [emergence] actually applies to anything', and approve of his philosophical etiquette in allowing this fact; but we finish our reading with everything in ruins, as Broad points to the formation of salt crystals as an example of this powerful and mysterious phenomenon of emergence in operation! Here he is [8, p.62] on the 'emergence' of water from hydrogen and oxygen:

Oxygen has certain properties and Hydrogen has certain other properties. They combine to form water, and the proportions in which they do this are fixed. Nothing that we know about Oxygen by itself or in its combinations with anything but Hydrogen would give us the least reason to suppose that it would combine with Hydrogen at all ... Here we have a clear instance of a case where, so far as we can tell, the properties of a whole composed of two constituents could not have

been predicted from a knowledge of the properties of these constituents taken separately . . .

Unfortunately, this was not an error peculiar to Broad. Part of the vitality of the emergentist movement at that time derived from such scientific examples. But within a very short time, increased understanding of how chemical reactions could be explained in terms of (reduced to) subatomic physics was to undermine this key component of emergentist thinking, and play a key role in the subsequent decline of the movement's influence. As McLaughlin puts it [44, pp. 89-90]:

In a span of roughly one hundred years, British Emergentism enjoyed a great rise and suffered a great fall. . . . It is one of my main contentions that advances in science, not philosophical criticism, led to the fall of British Emergentism . . . In their quest to discover "the connexion or lack of connexion of the various sciences" (Broad 1923, pp. 41-42), the emergentists left the dry land of the a priori to brave the sea of empirical fortune . . . and for a while winds of evidence were in their sails; but the winds gradually diminished, and eventually ceased altogether to blow their way. Without these winds in its sail, the British Emergentist movement has come to an almost complete halt.

Abstracting features of the real world can be a dangerous, even reckless, occupation. The abstractions need to be clearly defined, and applied to well-understood real contexts. Emergence should not be used too loosely as a repository for things we do not yet understand. On the other hand, we will see that the identification of true emergence arguably depends on an analysis of definability in the real world. And given the number of intractable problems the notion of definability throws up even in the 'safer' mathematical context, it is no surprise that mistakes will inevitably occur.

7 THE MATHEMATICS OF EMERGENCE

Things have not changed that much in the eighty years since the heyday of British emergentism. One knows exactly what Ronald Arkin is referring to when he says ([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?* (My emphasis)

The success of the Turing Test for machine intelligence inspired the devising of an observer-based Test for Emergence, by Ronald, Sipper and Capcarrère [62]. Here is an outline of the basic criteria they list for the validation of emergence

within a given system – where the the key player here, the observer, may also be the system designer:

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 .

There is a parallel between the Turing Test context and that of the Emergence Test, in that the role of observer is reflects a practical reality. On the other hand, the element of peer review, so appropriate to the task of recognising machine intelligence, does not fit the emergence-recognition task so well, even for emergence related to mental processes. What the test does do is clarify somewhat that practical reality mentioned above. A nice touch is the formal distinguishing of the difference between the languages pertaining to different levels of observation, preparing the way for mathematical analysis of relationships between these languages, and comparisons of relative complexities.

There are already a number of areas in which putative emergence has been approached via mathematical frameworks in which the theory necessarily delivers descriptions framed in explicit and concise language. What is usually apparent in such cases is the presence of non-linearity and an associated presence of mathematically generated infinities and of implicit quantifiers.

In the 1950s, Alan Turing followed through on an old interest in the origin of patterns in nature, proposing a simple reaction-diffusion system describing chemical reactions and diffusion to account for morphogenesis, i.e., the development of form and shape in biological systems. Some of his best-known work in this direction relates to phyllotaxis and the relationship of the standard Fibonacci series to the arrangement of leaves or petals etc. in plants. Typical of Turing, his work on this seems to combine detailed formalisation of concepts and working out of unexpected mathematical relationships in particular cases, with much more far-reaching and visionary ambitions. In a letter to the leading physiologist Professor J.Z. Young, dated 8th February, 1951, Turing talks of working on:

... my mathematical theory of embryology ... This is yielding to treatment, and it will so far as I can see, give satisfactory explanations of
i) Gastrulation.

- ii) Polygonally symmetrical structures, e.g., starfish, flowers.
- iii) Leaf arrangement, in particular the way the Fibonacci series (0,1,1,2,3,5,8,13,...) comes to be involved.
- iv) Colour patterns on animals, e.g., stripes, spots and dappling.
- v) Patterns on nearly spherical structures such as some Radiolara, but this is more difficult and doubtful.

I am really doing this now because it is yielding more easily to treatment. . . .

Turing goes on to mention connections with more difficult questions related to brain structure, leading Hodges to speculate that Turing's eventual aim was a logical description of the nervous system. It is this overall vision of drawing out the mathematical underpinnings of nature, ones basic enough to have application to the human brain, which is specially relevant to the topic of this article.

See the webpage on Alan Turing and Morphogenesis maintained by Jonathan Swinton [72] for a wealth of further information.

Under the *synergetics* label comes a whole body of mathematically framed approaches to emergence, associated with the physicist Hermann Haken. Another key figure in the area is Ilya Prigogine, whose work on the mathematical modelling of nonlinear and irreversible processes and dissipative structures brought him the 1977 Nobel Prize in Chemistry. Michael Bushev's book [11] *Synergetics - Chaos, Order, Self-Organization* gives a useful overview of the area.

While drawing technically on the analysis of phase-transitions in thermodynamics, synergetics has grown into an interdisciplinary field dealing with the origins and evolution of macroscopic patterns and spacio-temporal structures in interactive systems. The emphasis is on mapping out *self-organisational* processes in science and the humanities – e.g., autopoiesis.

Synergetics, with its focus on the notion of self-organisation, shifts the viewpoint on emergent phenomena. We switch from interested and surprised observer, to proactive participant groping for the descriptive framework in which to express the connection between microscopic and macroscopic structure. This is the perspective we will maintain throughout the remainder of this article.

8 DEFINABILITY AND EMERGENT STRUCTURE

For the most part, our relationship with emergence involves observation, surprise, and attempts to describe the link between a basic, well understood interactive local structure, and observed emergent patterns of behaviour and more advanced formations. For the most part we follow the familiar route from empirical data to theory — where even a two-way, dialectical interaction between the two does not disturb the essential dominance of the real world in the process. For instance, this is what Turing did for the role of Fibonacci numbers in relation to the sunflower etc.

Theoretically, the element of surprise may arise from the elegance or cleverness of the description discovered. And if the character of the description produces an emotional response in the mathematician or scientist, it is a reasonable guess that the close relationship between the material phenomenon and its description may mean that the character of the description has a key role in constraining that phenomenon in observably surprising ways. In mathematics, it is well-known that complicated descriptions may take us beyond what is computable — so we certainly have here a potential source of surprise in emergence.

We are now close to capturing an important intuition concerning the relationship between descriptions and the real world, whereby descriptions come to assume the dominant role; the intuition is that entities *exist because of, and according to*, mathematical laws. This is what Leibniz [40] wrote in *The Monadology*, section 32, explaining his *principle of sufficient reason*:

... by virtue of which there can be found no fact that is true or existent, or any true proposition, without there being a sufficient reason for its being so and not otherwise, although we cannot know these reasons in most cases.

We can think of ‘sufficient reason’ as being in a broad sense a *description* of how something arises from a given context, and ‘fact’ as something we observe to have a real-world validity or existence. It then follows that natural phenomena not only generate descriptions, but arise and derive form from them. The picture is of a world in which basic structure gives rise to further structure describable from below. We then have two important observations: One is that this is redolent of the emergentists’ layering of scientific fields; and secondly, that it connects with a useful abstraction — the concept of mathematical definability, which formalises what we mean by describability in a mathematical structure. The payoff is a new precision to our experience of emergence as a potentially non-algorithmic determinant of events.

Of course, definability is relatively undervalued as a useful mathematical concept. Unlike, say, ‘consistency’, which has long since been usefully assimilated into the culture, and is no longer recognised as coming out of the logicians toolbox — definability plays a fairly minor role in everyday discourse. Definability is a key concept for those trying to formulate theories about the real universe, with foundational work in this context going back to Hans Reichenbach [58] in 1924, and somewhat later, Carl Gustav Hempel [29]. The mathematical framework within which definability is discussed nowadays was formulated over a period of years, and first appeared in Alfred Tarski’s 1931 paper [73] (translated, and reprinted, in [74]).

When one considers definability in relation to how the universe provides ‘sufficient reason’ for the existence and uniqueness of its own component entities, one needs to deal with the fact that language is a human invention. On the other hand, the elements of logical structure underlying descriptions in a formal first-order language one is fairly safe in assuming to be features of any recognisable universe. It is basic causal relationships which we must consider to be essential ingredients

of any universe with enough regularities to be subjected to scientific analysis, and those from which logical structure is derived are arguably the simplest. This leads us to the reasonable assumption that descriptions, giving rise to observed structure in the universe, are not dependent on man-made language, quite the reverse.

There is a language-free approach to identifying those relations on a structure which are fixed uniquely by its basic relations, involving looking at the *automorphisms* of the structure. We say that a relation on a structure is *invariant* if and only if it remains fixed under all automorphisms of the structure. Under the foregoing assumptions about language capturing fundamentals of a structure, invariance is a generalisation of the notion of definability. This does highlight the fact that in a language-dependent expression of ‘sufficient reason’ in a real context, the particular character of the language used to frame descriptions is very important — in fact, it is well-known that invariance can be characterised as a definability notion in a suitably extended language.

Associating emergence with precise notions of definability immediately leads us to expect a higher level of robustness of emergent phenomena than might otherwise expect. Replicating an interactive environment would lead one to expect the emerging phenomena, captured via definable relations, to be somewhat the similar in the distinct but similar contexts. And this is very much what one finds in practice: this is an intuition supported by experience. Here is Martin Nowak (Director, Program for Evolutionary Dynamics, Harvard University) writing in John Brockman’s edited volume of distinguished scientists confessing to *What We Believe But Cannot Prove* [9]:

I believe the following aspects of evolution to be true, without knowing how to turn them into (respectable) research topics.

Important steps in evolution are robust. Multicellularity evolved at least ten times. There are several independent origins of eusociality. There were a number of lineages leading from primates to humans. If our ancestors had not evolved language, somebody else would have.

In the present context, our universe is that of the human brain. And here, as in the wider physical universe, we have a very good picture of the basic interactions within the brain. What we need is to compare our refined model of emergence with what we know about the brain and the higher mental functionality we would like to relate it to.

9 SUPERVENIENCE AND EMERGENCE IN FOCUS

According to Jaegwon Kim [34, pp. 14–15], supervenience:

...represents the idea that mentality is at bottom physically based, and that there is no free-floating mentality unanchored in the physical nature of objects and events in which it is manifested.

There are different ways of expressing this more formally, but the informal notion will serve us well enough for now.

The hope is for a clarified notion of emergence helping us pin down the nature of supervenience, and so, of intelligent thought and other aspects of human mentality. The aim is a model which reconciles the pressing and apparently inconsistent claims of physicalism and dualism on our intuitions about the mind. These post-Cartesian expectations include:

- Achieving a non-reductive physicalism, delivering –
- Mind-body supervenience;
- The physical irreducibility of the mental – including consciousness, qualia;
- And the causal efficaciousness of the mental;
- With definability removing conflict between vertical determination and horizontal causation.

But in doing this, we have a number of tasks. One is to identify a role for emergence in mental activity, given the warnings we encountered earlier concerning too loose an application of the criteria for emergence. Another is to examine the extent to which emergence equips our abstraction of mentality, and to isolate phenomena needing further explanation. And, having done this, to fill out, however schematically we allow ourselves to work, the mathematical model within which we want our appropriate notion of definability to be operative.

For our first task we should enlist the help of the Emergence Test. The basic design of the brain is by now well understood. The local interactivity has even been mathematically modelled, and we return to this below; on the other hand, there are mental phenomena which involve non-local activities, which could well point to emergence at work, but whose content, and qualifications for passing the surprise criterion 3 of the Emergence Test, are, in this context, definitely best assessed by us as observers of mentality. (The need for different modes of description for the different levels here — corresponding to languages \mathcal{L}_1 and \mathcal{L}_2 of the test — is not in question.)

A rich source of such observations is Jacques Hadamard's 1945 book [27] on *The Psychology of Invention in the Mathematical Field*. Hadamard's conversations with Poincaré, who also had a special interest in mental creativity, provided much of the material for the book. In mathematical thinking, there is clearer division than usual between the exercise of pure reason and of more creative thinking. A striking instance of this is seen in the following account of Hadamard:

At first Poincaré attacked [a problem] vainly for a fortnight, attempting to prove there could not be any such function ... [quoting Poincaré]:

‘Having reached Coutances, we entered an omnibus to go some place or other. At the moment when I put my foot on the step, the idea came to me, without anything in my former thoughts seeming to have paved the way for it ... I did not verify the idea ... I went on with a conversation already commenced, but I felt a perfect certainty. On my return to Caen, for conscience sake, I verified the result at my leisure.’

What is most striking about this anecdote is not so much the surprise ingredient, required by part 3 of the Emergence Test, although the hidden origin of Poincaré's solution to his problem is just what we are looking for in an emergent event. There is a *robustness* to Poincaré's unexpected idea, a 'perfect certainty'. He did not need to write anything down, he was able to carry on a conversation, and carry his valuable solution home entire, for later verification. This is just what we would expect from our refined model of emergence in terms of definability. Definability had given the idea a memetic quality.

Just to be sure this was not just a one-off inexplicably bizarre incident, here is Hadamard again quoting Poincaré:

'Then I turned my attention to the study of some arithmetical questions apparently without much success ... Disgusted with my failure, I went to spend a few days at the seaside and thought of something else. One morning, walking on the bluff, the idea came to me, with just the same characteristics of brevity, suddenness and immediate certainty, that the arithmetic transformations of indefinite ternary quadratic forms were identical with those of non-Euclidian geometry.'

Of course, most mathematicians have had some such experience, even if not so often as Poincaré. Alan Turing seems to have had something like this in mind when he gave the following interpretation of his 1939 attempt [77] to transcend the theoretical constraints presented by Gödel's incompleteness theorem:

Mathematical reasoning may be regarded ... as the exercise of a combination of ... intuition and ingenuity ... In pre-Gödel times it was thought by some that all the intuitive judgements of mathematics could be replaced by a finite number of ... rules. The necessity for intuition would then be entirely eliminated. In our discussions, however, we have gone to the opposite extreme and eliminated not intuition but ingenuity, and this in spite of the fact that our aim has been in much the same direction.

Logicians will know that the device Turing had tried to use to go beyond Gödel was a *representational* one, using recursive ordinals; and that the obstacle encountered was the inaccessibility of particular ordinals to algorithmic grasp.

10 IS THAT ALL THERE IS? - REPRESENTATION

Now let's return to the question of design featuring in parts 1 and 2 of the Emergence Test. How exactly do we bridge the gap between 'emergent' higher mental functionality and ... *what* algorithmic 'design'? Answering is difficult. As Rodney Brooks points out in *Nature* in 2001:

... neither AI nor Alife has produced artifacts that could be confused with a living organism for more than an instant.

The alarm bells ringing do not just concern getting the right underlying model. The modellers have done their best and met with limited success, making us ask if emergence is sufficient to explain what we observe in intelligent thought . . . *is that all there is?*

And what about connectionist models of computation? They certainly seem to promise exciting things, including the delivery of more than a Turing machine does, as observed by 2005 David E. Rumelhart Prize recipient, Paul Smolensky, in an article [69] in 1988:

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.

Obviously, connectionist models have come a long way since Turing's discussion [78] of 'unorganised machines' — see Christof Teuscher's fascinating book [75] on *Turing's Connectionism* for a full account of how Turing "was the first person to consider building artificial computing machines out of simple, neuron-like elements connected together into networks in a largely random manner" — and the McCulloch and Pitts [43] early paper on neural nets.¹ Despite disputes about the adequacies of neural network model — the dissenting work of Minsky and Papert [45] had a big effect on the area — the approach is still responsible for numerous innovative and exciting developments (see the by now classic contribution of McClelland and Rumelhart [42]).

But for Steven Pinker "...neural networks alone cannot do the job". And focussing on our elusive higher functionality, Pinker points to a "kind of mental fecundity called recursion". By this he seems to have in mind the re-use of *emergent* mental images within mental processes which are responsible for such emergence. This clearly goes beyond the sort of standard recursions one can commonly find neural nets performing. Here is an illustrative example from Pinker [50]:

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.

The new element in such recursions is *representation* of non-local collections of data, and again one can identify definability as providing a key concept here.

¹It is an interesting observation on the history of computability after Turing, that even Stephen Kleene, sometimes held responsible for computability (or 'recursive function theory' as he called it) becoming a niche technical area with little connection with the real world, was himself interested in McCulloch and Pitts' paper. Out of this interest grew his own 1956 paper [36] on neural nets (a paper which one will search for in vain in the standard texts on recursion theory).

The non-locality is something familiar to neuroscientists, as is the reaching for a physical mechanism for implementing that representation process which feeds into the recursions apparent at the mental level. Making a similar point in the following passage [16, p. 170] from his book on *The Feeling Of What Happens*, the neurologist Antonio Damasio gives a nice description of the hierarchical development of a particular instance of consciousness within the brain, 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.

The picture is one of 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. The key conception is that of computational loops incorporating, in a controlled way, these 'second-order' aspects of the computation itself. This reminds us of the layering scenario C.D. Broad, but there is something more complicated going on.

11 DEFINABILITY IN WHAT STRUCTURE?

There are, in addition to neural nets, other computational models expressing metaphors for natural processes. Examples include quantum and molecular computing, membrane computing, cellular automata, L-systems, DNA computing, swarm and evolutionary computation, and relativistic computing. Many of these have basic algorithmic content which gives rise to new emergent forms, but none of them seem to have the power shown by the human mind to represent these forms in such a way as to convert them into fodder for their own algorithmic appetites. Here is Damasio again [16, p. 170]:

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 relationship 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.*

In order to get at the essence of what is happening here, one needs to get at the essential computational content of definability. One needs to strip away the particularities of diverse natural contexts, and abstract from nature a model which will reveal to us the theoretical constraints on emergence and definability — in the same way that Gödel, working with formal systems, revealed to us basic constraints on algorithmically accessed knowledge. The outcome in prospect is not just a schematic understanding of mental processes, but a model of very general applicability.

Models based closely on particular aspects of nature may deliver very important practical computational benefits, even though there is always the difficulty of developing a *range* of algorithms implementable within a particular model. The benefits of abstraction are well-known. It facilitates depth of theory. It can put apparently diverse contexts within a common framework, enabling one to get an overview. This is important when one is trying to suitably locate emergent phenomena, in that the ‘special sciences’ which the early emergentists sought to arrange hierarchically do impinge on each other. One needs a model in which emergent entities can be both differentiated, and at the same time can keep their essential character as information transferable across scientific fields. It is true that the division into ‘special sciences’ is an empirical reality, and makes a lot of sense from the point of view of providing work-spaces within which Thomas Kuhn’s ‘normal science’ (the bread and butter of scientific research) can take place. But given the complexity of the brain and its mental attributes, in which the sort of recursions highlighted by Pinker and Damasio feature, there is a need for a corresponding level of theory which is not yet current. Such a theory needs to capture the abundant descriptive material relating many different contexts, but give it a precision and technical content which will clarify many of the current confusions and controversies. It may not provide us with the sort of practically useful reductions which people are commonly drawn to, but it may provide us with a better idea of the value and limitations of such reductions.

Looking to lessons from other areas, there are a number of fundamental problems facing physicists, for example. A number of these can be expressed as definability problems. High on the list of such problems relates to the standard model of particle physics. As Peter Woit describes it in the introduction to his book *Not Even Wrong: The Failure of String Theory and the Continuing Challenge to Unify the Laws of Physics*, [?]:

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

Later in the book, he explains: “One way of thinking about what is unsatisfactory about the standard model is that it leaves seventeen non-trivial numbers still to be explained, ...”. How does the universe *define* its own basic components, its laws

and entities? There are a number of more-or-less unsatisfactory ways of skirting round this question, in the main appealing to outside assistance from God or Many Worlds (with the addition of the anthropic principle), but the best answer would be an explanation of how the material universe *does it for itself*. Lee Smolin in his recent book *The Trouble With Physics* [70] lists 'Five Great Problems in Theoretical Physics', all of which can be related to basic questions of how and why certain observable features of the real universe arise, or, from our perspective, how the universe *defines* itself. Or, equally important, how it fails to define things. Here are two of Smolin's problems, slightly paraphrased:

- Combine general relativity and quantum theory into a single theory that can claim to be the complete theory of nature.
- Resolve the problems in the foundations of quantum mechanics

For the former, definability considerations in a suitable model might lead one to a negative answer: No such theory exists. And for the latter, allowing that the universe might *fail* to define entities uniquely a basic levels might clarify a lot.

But to return to our question: Definability in what structure? For Smolin, it is causality which is fundamental, and one has to admit that it is hard to say anything about a universe without causal relations. Or to put it another way, if one were to design a universe from nothing, and do it in as simple and incremental way as possible, one would need to have in mind a blueprint in which basic entities, involving simple mathematical relations, were stipulated. This how Smolin [70, p. 241] argues the case:

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 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.

So, to paraphrase Smolin, not only do global features of the universe constrain its causal structure; it is also the case that causal relations have comprise a structure which may determine the nature of the global relations on that structure. We are back with the basic mathematical notions of definability and invariance needed to express such notions. What Smolin has in mind is a range of approaches, associated with 'early champions of causality' such as Roger Penrose, Rafael Sorkin, Fay Dowker, and Fotini Markopoulou.

Of course, the notion of causality has been discussed since at least the time of Aristotle, while it was David Hume (in [31], for example) who most memorably brought out the need for more clarity. For us, causal relations are not just observed successions of contiguous events inductively extended; they involve some necessary

connection, which, since the time of Newton, we have expected to be mathematical in nature. Even Hume [31, p. 56] asserts:

Shall we rest contented with these two relations of contiguity and succession? By no means . . . there is a NECESSARY CONNEXION to be taken into consideration.

There are now two important observations concerning the mathematical modelling of causal relationships. The first is the obvious one that ever since Newton, we have tried, and to a large extent succeeded, in characterising *basic* causal relations in terms of functions over the real numbers (sometimes this real domain being implicit, and needing some coding to coax out); and that these functions tend to have a high degree of continuity and accessibility to computer processing. And the second observation is that there are other causal relations which are still within the scientific domain, and describable in terms of real numbers, but which do not appear to produce *computable* relationships. We saw examples of the latter earlier.

The scenario we have been building is one in which relations in science are those subject to descriptions, and, in fact, in terms of real numbers. And that these relations are subject to hierarchical classification in terms of their logical complexity in some appropriate formal language. The claim is that in each area of science, the basic relations are computable ones; that the limitation to such relations still gives us a very good correspondence between science and ‘reality’; and that these basic relations are truly basic in that the more complicated causality we encounter is somehow associated with — is in fact definable in terms of — our basic computable causal relations. Maybe the simplest example is Kreisel’s three-body situation. One can try and reduce the problem of predicting the motion to the computable two-body situation, but then the underlying infinitary nature of the non-linear mathematics produces something closely akin to an emergent phenomenon. The physical context is not obviously ‘global’ in the sense that the large-scale formations in the Cat’s Eye Nebula are, but there is a mathematical dimension to the causality which serves a similar role in making description more difficult, and prediction not computable.

In the next section we give a brief introduction to the Turing model of computation with real inputs, designed to capture the basic causal relations of Newton and science more generally. And the final section will look at what we can extract from what we know so far about about Turing invariance and definability.

12 THE TURING MODEL EXTENDED

The oracle Turing machine, which made its first appearance in Turing [77], should be familiar enough. The details are not important, but can be found in most reasonable introductions to computability (see for instance [13]).

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.

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. Depending on one's viewpoint, this is either a rather reduced scientific universe, or a much expanded one. The familiar mathematical presentation of it is due to Emil Post [53], in his search for the informational underpinnings of computational structure.

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*.

Schematically, as we have argued, any causal context framed in terms everyday computable mathematics can be modelled in terms of Turing reductions. In particular, viewing scientific observations as causal events, one can model these computations with oracles, as

Anyway, emergence can now be formalised as definability over the appropriate substructure of the Turing universe; or more generally, as invariance under automorphisms of the Turing universe. Any computable causal relation will be found amongst those listed by Turing and forming part of the Turing universe.

This brings us to a well-known research programme, initiated by Hartley Rogers in his 1967 paper [60], 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 in some disarray. 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 [38] on the structure of scientific revolutions, such tensions can be seen as signs that something scientifically important is at stake.

There are other computational reducibilities with good claims to model particular causal contexts. Particularly important is the generalisation of Turing reducibility in which non-deterministic computations are permitted. This provides a natural model of environments where computations call on auxiliary information which is possibly incomplete, so that an oracle would be forced to remain silent, possibly stalling the computation permanently while the machine waited for a non-existent reply from the oracle. In this situation, 'guessing' at certain points in a computation might enable one to avoid using information one cannot access. There are everyday examples of such situations. Mathematically (see [13, chap. 11]), the model is usually presented in terms of *enumeration reducibility*.

The bi-interpretability conjecture can be applied to other models of relative computation, but is quite sensitive to the way information is accessed. Turing reducibility and enumeration reducibility, which have strong claims to having captured the most general mode of effective computation from real information, including that involved in basic natural laws, human observation, and currently feasible computer functionality, seem to have a peculiarly complex theory. For them, there is a rich body of definable relations (as one would expect in a candidate model for aspects of real-world complexity), but relations definable in everyday language (by which we mean here second-order arithmetic) may still turn out to be undefinable in these models. Constrain the information-gathering capacity of the reducibility, and definability in the resulting structure becomes less interesting – for instance, in the structure of the *many-one* degrees, the zero degree consisting of the outright computable objects is the only one definable. But, moving towards the other end of the spectrum, if we allow a computation to call on (uncoded) infinitary information before completion, and one may get bi-interpretability, ruling it out as a route to modelling of the sort of ambiguity one sees at the quantum level, or in mental activity at the subconscious level. For instance, the hyperdegrees admit bi-interpretability.

13 MENTAL CAUSATION REVISITED

The test of any mathematical model is its success in providing persuasive explanations for natural phenomena for which previously there were at best conflicting interpretations. For better understanding of the relationship between brain and mentality, we still have the basic questions:

- How can mentality have a causal role in a world that is fundamentally physical?
- And what about overdetermination – the problem of phenomena having both mental and physical causes?

As Jaegwon Kim [35] concisely puts it:

... the problem of mental causation is solvable only if mentality is physically reducible; however, phenomenal consciousness resists physical reduction, putting its causal efficacy in peril.

What we have outlined above is how we can give computational content to emergence, use emergence to link functions of the brain to higher mental activity, use definability to contribute substance to recursions involving mental representations, and hence present the whole organism with a new coherence and organic entity.

Let us summarise what we now have at hand.

Firstly, it is clear that the notion of causality can be a misleading concept in this context. This is certainly true in relation to overdetermination — that is, the problem of mental phenomena having both physical and mental causes. As we have seen, causality in the everyday sense can be substantiated through the existence of mathematical relationships between events, which provide a definite mathematical connection from the caused to the causing. But when one looks at emergent phenomena, one can have both emergent objects *and* emergent relationships. The emergent relationships, at the newly emergent level, may be computable in the classical sense, whilst the relationship between the emergent phenomenon and the substratum on which it rests may be more complex, to the point of being beyond computation. There is clearly a causal relationship between emergent phenomenon and substratum, and we may have a definition of it, so that the mathematical qualification for vertical causality exists. But the computable relations at the emergent level may be much more in evidence, to the point where they become the basis for a scientific field, with its own mathematical relations and attending causality. Where is the *real* causality here? We need to recognise that the old notions are unhelpful, and that a full description of the mathematics involved is the best we can do. Further, we need to recognise that the different levels can interact in quite complex ways, and that some of these interactions, connecting emergent and basic levels can be quite basic and mechanistic themselves. This, of course, is the basis for the success of modern psychiatry, employing mechanistic interventions to modify the emergent mentality. What we are seeing here is a renewed dualism, but a more mature version of it. This is implicit in a recent paper [15] of Antonella Corradini, which has had an influence on the development of this article.

Kim comments [35, p. 15]:

Mind-body supervenience has been embraced by some philosophers as an attractive option because it has seemed to them a possible way of protecting the autonomy of the mental domain without lapsing back into antiphysicalist dualism.

But goes on:

how [is] it ... possible, on such a picture, for mentality to have causal powers, powers to influence the course of natural events.

This forms part of a section [35, pp. 13–22] on “The Supervenience/Exclusion Argument”, which can be used as an exercise (see particularly p. 19) in applying our more coherent conception of causality.

So – recognisable ‘causality’ occurs at different levels of the model, connected by relative definability. But this causality may occur in the form of relations with identifiable algorithmic content, this content at higher levels being emergent. The diverse levels form a unity, with the ‘causal’ structure observed at one level reflected at other levels — with the possibility of both computable and incomputable ‘feedback’ between levels.

One striking feature of the current debate about supervenience and non-reductive physicalism, is the extensive (and often very perceptive) contribution from those with a theological interest. The aim of setting the human mind free from the scientific cage of physicalism is a demanding and worthy project which melds well with a religious agenda. This is William Hasker, writing in *The Emergent Self* [28, p. 175]:

The “levels” involved are levels of organisation and integration, and the downward influence means that the behavior of “lower” levels – that is, of the components of which the “higher-level” structure consists – is different than it would otherwise be, because of the influence of the new property that emerges in consequence of the higher-level organization.

This is consistent with our model. Our model does preserve determinism through all levels, but is not Laplacian. There is no predictive demon permitted, there is no limit to the potential richness of information involved. The mathematical blueprint for mental attributes and activities may well be contained within the actuality of the physical brain. But cut open the brain, or subject it to less intrusive inspection, and you will see only imperfect evidence of the content of these levels. “We” are participants according to the newly emergent rules and entities, and our mentality is part of the organism. And there is no omniscient god even consistent with this model; we have at best a “self-evolving god”.²

Although the mathematical model discussed here might not be to the total satisfaction of, say, Samuel Alexander (whose emergent phenomena were held to be of quite mysterious origin), it does go some way towards rescuing consciousness, qualia and free will from the reductionists’ dead hand. For instance, mental causality is given the sort of status that the person in the street would accept, along with consciousness and qualia. These do ‘exist’. And the ubiquitous incomputability consequent on logically complex definitions raises the sort of everyday judgements we make above the level of what a computer does. There really can be adventure in thinking, involving hidden agendas in which the whole organism participates, and which certainly gives the feeling of free will, and allows us to live with a very valid assumption of the importance of human mental activity.

²A description due to my wife, I should say.

References

- [1] S. Alexander. *Space, Time, and Deity* Vol. 2, 1927.
- [2] R.C. Arkin. *Behaviour-Based Robotics*. MIT Press, 1998.
- [3] E.J. Beggs and J.V. Tucker. *Experimental computation of real numbers by Newtonian machines*. Proceedings Royal Society Series A, **463** (2007), 1541–1561.
- [4] E.J. Beggs, J.F. Costa, B. Loff and J.V. Tucker. *Computational complexity with experiments as oracles*. Proceedings Royal Society Series A, **464** (2008), 2777–2801.
- [5] L. Blum, F. Cucker, M. Shub and S. Smale. *Complexity and Real Computation*. Springer, 1997.
- [6] L. Blum and S. Smale. *The Godel incompleteness theorem and decidability over a ring*. In *From Topology to Computation: Proceedings of the Smalefest (Berkeley, CA, 1990)*, Springer, New York, 1993, pp.321–339.
- [7] V. Brattka. *The emperors new recursiveness: the epigraph of the exponential function in two models of computability*. In *Words, languages & combinatorics, III (Kyoto, 2000)*. World Sci. Publishing, River Edge, NJ, 2003, pp.63–72.
- [8] C.D. Broad. *The Mind and Its Place In Nature*. Kegan-Paul, London, 1923.
- [9] J. Brockman (ed.). *What We Believe but Cannot Prove: Today's Leading Thinkers on Science in the Age of Certainty*, Harper Perennial, New York, 2006.
- [10] R. A. Brooks. *Elephants Dont Play Chess*. Robotics and Autonomous Systems, **6** (1990), 3–15.
- [11] M. Bushev. *Synergetics – Chaos, Order, Self-Organization*. World Scientific, 1994.
- [12] 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 Note Series 259, Cambridge University Press, Cambridge, 1999, pp. 63–116.
- [13] S.B. Cooper. *Computability Theory*. Chapman & Hall/CRC, Boca Raton, London, New York, Washington, D.C., 2004.
- [14] S.B. Cooper and P. Odifreddi. *Incomputability in Nature*. In S.B. Cooper and S.S. Goncharov. *Computability and Models*. Kluwer Academic/Plenum, New York, Boston, Dordrecht, London, Moscow, 2003, pages 137–160.

- [15] A. Corradini. *The emergence of mind. A dualistic understanding*. In A. Carsetti (ed.), *Causality, Meaningful Complexity and Knowledge Construction*, Springer, 2009.
- [16] A.R. Damasio. *The Feeling Of What Happens: Body and Emotion in the Making of Consciousness*, Harcourt Brace, 1999.
- [17] D. Davidson. *Mental Events*. In *Experience and Theory* (Foster and Swanson, eds.), Duckworth, London, 1970. Reprinted in Davidson, 2001.
- [18] D. Davidson. *Essays on Actions and Events*, 2nd ed. Oxford University Press, Oxford, 2001.
- [19] M. Davis. *The Universal Computer: The Road from Leibniz to Turing*. W.W. Norton, New York, London, 2000.
- [20] M. Davis. *The myth of hypercomputation*. In *Alan Turing: Life and legacy of a great thinker* (C. Teuscher, ed.), Springer-Verlag, Berlin, Heidelberg, 2004, pp. 195–212.
- [21] D.C. Dennett. *Consciousness Explained*. Little, Brown & Company, 1991.
- [22] R. Descartes. *Discourse on Method*. Paris, 1637.
- [23] D. Deutsch. *Quantum theory, the Church Turing principle, and the universal quantum computer*. Proc. Roy. Soc., A 400 (1985), 97–117.
- [24] F. Engels. *Socialism: Utopian and Scientific*. In *Revue Socialiste*, 1880. Reprinted in *Marx/Engels Selected Works* (trans. from the French by Paul Lafargue in 1892), vol. 3, pp.95–151.
- [25] Yu.L Ershov, S.S. Goncharov, A. Nerode, J.B. Remmel (Editors). *Handbook of Recursive Mathematics*, Volumes 1 and 2. Elsevier, Amsterdam, New York, Oxford, Tokyo, 1998.
- [26] E.R. Griffor (Editor). *Handbook of Computability Theory*. Elsevier, Amsterdam, New York, Oxford, Tokyo, 1999.
- [27] J. Hadamard. *The Psychology of Invention in the Mathematical Field*. Princeton Univ. Press, Princeton, 1945.
- [28] W. Hasker. *The Emergent Self*. Cornell University Press, Ithaca, London, 1999.
- [29] C.G. Hempel. *Fundamentals of Concept Formation in Empirical Science*. University of Chicago Press, Chicago, 1952.
- [30] A. Hodges. *Alan Turing: The Enigma*. Vintage, London, Melbourne, Johannesburg, 1992.
- [31] D. Hume. *A Treatise of Human Nature* (1740). Reprinted by Oxford University Press, Oxford, 1967.

- [32] T.H. Huxley. *On the Hypothesis that Animals are Automata, and its History*. *The Fortnightly Review*, n.s. **16** (1874), 555–580. Reprinted in *Method and Results: Essays by Thomas H. Huxley*, D. Appleton and Company, New York, 1898.
- [33] J. Kim. *Mind and Supervenience*. Cambridge University Press, Cambridge, 1993.
- [34] J. Kim. *Mind in a Physical World*. MIT Press, 1998.
- [35] J. Kim. *Physicalism, or Something Near Enough*. Princeton University Press, Princeton, Oxford, 2005.
- [36] S.C. Kleene. *Representation of events in nerve nets and finite automata*. In C. Shannon and J. McCarthy (eds.), *Automata Studies*, Annals of mathematics studies, no. 34, Princeton University Press, Princeton, N.J., 1956, pp. 3–42.
- [37] 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.
- [38] T.S. Kuhn. *The Structure of Scientific Revolutions*. Third edition 1996, University of Chicago Press, Chicago, London.
- [39] P.S. de Laplace. *Essai Philosophique sur les Probabilités*. English trans. by F.W. Truscott and F.L. Emory, Dover, New York, 1951.
- [40] G.W. Leibniz. *La Monadologie* (1714). English translation by G.M. Ross, 1999.
- [41] Y. Matiyasevich. *Hilbert's Tenth Problem*. MIT Press, Cambridge, Mass., London, 1993.
- [42] J.L. McClelland and D.E. Rumelhart. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press, Cambridge, MA, 1986.
- [43] W. McCulloch and W. Pitts. *A logical calculus of the ideas immanent in nervous activity*. *Bull. Math. Biophys.* **5** (1943), 115–133.
- [44] B.P. McLaughlin. *The Rise and Fall of British Emergentism*. In *Emergence or Reduction? – Essays on the Prospects of Nonreductive Physicalism* (A. Beckermann, H. Flohr, J. Kim, eds.), de Gruyter, Berlin, 1992, pp. 49–93.
- [45] M. Minsky and S. Papert. *Perceptrons: An Introduction to Computational Geometry*, MIT Press, Cambridge MA, 1969.
- [46] P. Odifreddi. *Classical Recursion Theory*. North-Holland/Elsevier, Amsterdam, New York, Oxford, Tokyo, 1989.

- [47] P. Odifreddi. *Kreisel's Church*. In *Kreiseliana: About and Around Georg Kreisel* (ed. P. Odifreddi). A K Peters, 1996. Reprinted in Olszewski, Wolenski and Janusz, pp.353–382.
- [48] A. Olszewski, J. Wolenski and R. Janusz (eds.). *Church's Thesis After 70 Years*. Ontos Verlag, 2006.
- [49] R. Penrose. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. Oxford University Press, Oxford, New York, Melbourne, 2002.
- [50] S. Pinker. *How the Mind Works*. W.W. Norton, New York, 1997.
- [51] G.J. Popek and R.P. Goldberg. *Formal requirements for virtualizable third generation architectures*. Communications of the ACM **17** (1974), 412–421.
- [52] E.L. Post. *Finite combinatory processes – Formulation 1*. J. of Symbolic Logic **1** (1936), 103–105; reprinted in Post (1994).
- [53] E. L. Post. *Degrees of recursive unsolvability: preliminary report*. (abstract). Bull. Amer. Math. Soc. **54** (1948), 641–642.
- [54] E.L. Post. *Solvability, Provability, Definability: The Collected Works of Emil L. Post* (Martin Davis, Editor). Birkhäuser, Boston, Basel, Berlin, 1994.
- [55] H. Putnam. *Minds and Machines*, reprinted in Putnam (1975).
- [56] H. Putnam. *The Nature of Mental States*, reprinted in Putnam (1975).
- [57] H. Putnam. *Mind, Language, and Reality*. Cambridge University Press, Cambridge, 1975.
- [58] H. Reichenbach. *Axiomatik der relativistischen Raum-Zeit-Lehre*, 1924. English translation: *Axiomatization of the theory of relativity*. University of California Press, 1969.
- [59] C. Reid. *Hilbert*. Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, 1986 (paperback edition).
- [60] 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.
- [61] H. Rogers, Jr. *Theory of Recursive Functions and Effective Computability* (1967). Reprinted by MIT Press, Cambridge, Mass., London, 1987.
- [62] E.M.A. Ronald, M. Sipper and M.S. Capcarrère. *Design, observation, surprise! A test of emergence*. Artificial Life, **5** (1999), 225–239.
- [63] G. Ryle. *The Concept of Mind*. University of Chicago Press, Chicago, 1949.

- [64] D. G. Saari and Z. Xia. *Off to Infinity in Finite Time*. Notices of the Amer. Math. Soc. **42** (1995), 538–546.
- [65] J.R. Searle. *The Mystery of Consciousness*. Granta, London, 1997.
- [66] R. Shaw. *Strange attractors, chaotic behaviour, and information flow*. Z. Naturforsch, **36A** (1981), 80–112.
- [67] R. Shaw, *The dripping faucet as a model chaotic system*, The Science Frontier Express Series, Aerial Press, Santa Cruz, CA, 1984.
- [68] A. Sloman, *Some requirements for human-like robots: Why the recent over-emphasis on embodiment has held up progress*. In : Creating Brain-Like Intelligence: From Basic Principles to Complex Intelligent Systems (B. Sendhoff, E. Körner, O. Sporns, H. Ritter, K. Doya, Eds.). Lecture Notes in Computer Science Vol. 5436, Springer 2009, pp. 248–277.
- [69] P. Smolensky. *On the proper treatment of connectionism*. Behavioral and Brain Sciences **11** (1988), 1–74.
- [70] L. Smolin. *The Trouble With Physics: The Rise of String Theory, the Fall of a Science and What Comes Next*. Houghton Mifflin, 2006.
- [71] R.I. Soare. *Recursively Enumerable Sets and Degrees*. Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, 1987.
- [72] J. Swinton. *Alan Turing and morphogenesis*. Part of Jonathan Swinton’s homepage: <http://www.swintons.net/jonathan/turing.htm>.
- [73] A. Tarski. *Sur les ensembles définissables de nombres réels. I*. Fundamenta Mathematicae, 17 (1931), 210–239.
- [74] A. Tarski. *On Definable Sets of Real Numbers*, translation of Tarski (1931) by J.H. Woodger. In A. Tarski, *Logic, Semantics, Metamathematics*, second edition, ed. by J. Corcoran. Hackett, Indianapolis, 1983, pp. 110–142.
- [75] C. Teuscher. *Turing’s Connectionism: An Investigation of Neural Network Architectures*. Springer-Verlag, London, Berlin, Heidelberg, 2002.
- [76] Turing A. M. (1936) *On computable numbers, with an application to the Entscheidungsproblem*. Proc. London Math. Soc. (2) **42** (1936–7), pp. 230–265. Reprinted in A. M. Turing, *Collected Works: Mathematical Logic*, pp. 18–53.
- [77] A.M. Turing. *Systems of logic based on ordinals*. Proc. London Math. Soc. (2) **45** (1939), pp. 161–228. Reprinted in A. M. Turing, *Collected Works: Mathematical Logic*, pp. 81–148.

- [78] A. M. Turing. *Intelligent machinery*. National Physical Laboratory Report (1948). In *Machine Intelligence 5* (B. Meltzer and D. Michie, eds.), Edinburgh University Press, Edinburgh, 1969, pp. 3–23. Reprinted in A. M. Turing, *Collected Works: Mechanical Intelligence* (D. C. Ince, ed.), North-Holland, Amsterdam, New York, Oxford, Tokyo, 1992, pp. 107–127.
- [79] A.M. Turing. *Computing machinery and intelligence*. *Mind* **59** (1950), 433–460. Reprinted in A. M. Turing, *Collected Works: Mechanical Intelligence* (D. C. Ince, ed.), North-Holland, Amsterdam, New York, Oxford, Tokyo, 1992.
- [80] A.M. Turing. *Collected Works: Mathematical Logic* (R.O. Gandy and C.E.M. Yates, Editors). Elsevier, Amsterdam, New York, Oxford, Tokyo, 2001.
- [81] J. van Leeuwen and J. Wiedermann. *The Turing Machine Paradigm in Contemporary Computing*. In *Mathematics Unlimited – 2001 and Beyond* (eds. B. Enquist and W. Schmidt), LNCS, Springer, 2000.
- [82] M. White. *Isaac Newton – The Last Sorcerer*. Fourth Estate, London, 1997.