TURING PAPERS, VOLUME I

The Turing Test, Turing Machines and the Church-Turing Thesis

edited by Peter Millican and Andy Clark

Introduction

Peter Millican, University of Leeds

This is the first of two volumes of essays in honour of Alan Turing. It is centred on the continuing discussion of his classic contributions to the theory of Artificial Intelligence and Computer Science, and in particular, the three most fundamental and seminal ideas universally associated with his name: the Turing Test, the Turing Machine and the Church-Turing Thesis.

The Turing Test was first proposed in a paper that was to become amongst philosophers (or at least those not specialising in logic and computation) Turing's easily best known work: "Computing Machinery and Intelligence", published in *Mind* in 1950. It was the fortieth anniversary of this publication that brought together, at the Turing 1990 Colloquium, the impressive interdisciplinary array of speakers and distinguished invited guests whose contributions, both at the Colloquium itself and subsequently, form the heart of these two volumes. The level of discussion at the Colloquium was such as to lead a number of the contributors to wish substantially to revise or extend their papers, and given the significance of the occasion, the relatively timeless nature of much of the subject matter, and the unusual opportunity for mutual response between researchers across a variety of disciplines, it seemed

appropriate to delay immediate publication for this purpose. We hope that this has allowed these collections, though conceived at Turing 1990, to be more than just another conference proceedings, albeit one with an unusually impressive cast of contributors. Amongst that cast, it is particularly gratifying to be able to include two of the earliest and most influential pioneers of Artificial Intelligence, Donald Michie (who worked side by side with Turing, codebreaking at Bletchley Park) and the Nobel laureate Herbert Simon, and also Turing's doctoral student Robin Gandy, who probably knew him better than anyone else now living.

The Turing Test and the Imitation of Human Cognition

Most of the papers in this volume allude in some way to the famous Turing Test, or the "imitation game" on which it is based, and several make it a central theme. The idea is very well known: an interrogator is connected by teletype to two respondents, one of which is a human and the other a computer programmed to respond like a human. The interrogator then asks questions of each respondent, with a view to discovering which of them is which. Turing argues that such an imitation game provides a useful criterion of intelligence - if the computer were able to give sufficiently human-like responses to resist identification in such circumstances, then it would be quite gratuitous to deny that it was behaving intelligently, irrespective of its alleged lack of a soul, an inner perspective, consciousness, or whatever.

Robert French does not dispute the Turing Test's adequacy as a positive criterion of intelligence, but he casts doubt on its usefulness by denying that any computer, intelligent or otherwise, could reasonably be expected to pass it. The problem he identifies is that the responses of any human in the imitation game will manifest numerous "subcognitive" influences which will be virtually impossible for any respondent to mimic unless it has experienced the world in a human-like way - the richness of our experience sets up a vast array of subcognitive associations which cannot realistically be formulated explicitly, or represented in a program, but which can be elicited by appropriately designed questions (e.g. "Rate 'Flugly' as the name of a teddy bear, and as the name of a glamorous female movie star"). The upshot is that the Turing Test is too demanding: it "provides a guarantee not of intelligence but of culturally-oriented *human* intelligence".

Donald Michie's paper further brings out the significance of the subconscious levels of human thought, though he suggests that French's attack on the Turing Test may be slightly uncharitable given

that Turing himself set a relatively low hurdle of success in the imitation game - the ability to deceive only to the extent of giving “an average interrogator” no more than a “70 per cent chance of making the right identification after five minutes of questioning” (it is also worth bearing in mind here the comments that Gandy makes regarding Turing’s intentions in his famous paper). Michie begins his own discussion by drawing attention to a hitherto little-known 1947 lecture in which Turing proposed a number of trail-blazing ideas which later became quite standard in computer science. Particularly relevant here is the notion that advanced computers would need a learning capacity, and in particular, the ability to learn from contact with human beings and to “adapt to their standards”. Michie points out that learning from humans is far from straightforward - typically as humans acquire expertise in a particular domain they become progressively *less* rather than more able to articulate their knowledge, and so effective learning from human experts must in practice involve the computer not just in passive reception of the expert’s opinions, but in active induction of the rules being implicitly followed. This reveals a surprising flaw in the Turing Test - questions aimed at eliciting explicit knowledge of these rules could be answered much better by such a “superarticulate” computer than by the human expert, but it seems odd to judge the machine a failure in the Test when it betrays its non-humanity by a superior cognitive performance.

Blay Whitby agrees with the negative points made by French and Michie regarding the adequacy of the Turing Test when interpreted as an operational definition of intelligence. However he regards this as a serious misinterpretation with unfortunate consequences, suggesting that it has led researchers in Artificial Intelligence to put far too much emphasis on the imitation of human performance, rather than on the achievement of a proper understanding of the abstract nature of intelligence, and on autonomous developments based on such an understanding. He argues that the imitation game can instead more usefully be seen as a persuasive device to encourage a paradigm shift towards the now familiar perspective which distinguishes between the physical and the logical nature of a machine, and which regards these as potentially independent. The game forces the interrogator to judge respondents on the basis of their input/output behaviour rather than their physical characteristics, and it is this shift of perspective, together with a related emphasis on the significance of third-person attitudes rather than intrinsic characteristics in the ascription of “intelligence”, that marks the proper legacy of Turing’s paper.

Ajit Narayanan takes up this theme of the importance of third-person attitudes in the ascription of intelligence, and begins from a reinterpretation of the imitation game based on Daniel Dennett’s idea of the “intentional stance”. According to this, the appropriate issue becomes not whether a computer

can think, but rather whether it can properly have the intentional stance applied to it - whether its behaviour can usefully be seen as intentionally directed. This, however, is clearly an insufficient condition for consciousness or for any other more full-blooded notion on Dennett's "ladder of personhood", so Narayanan proceeds to formulate a distinction between the "representational stance" (concerned with the applicability of terms from a representational framework, based on behaviour alone) and the "ascriptional stance" (which requires in addition a commitment to some underlying theory of consciousness, considered to be appropriate to the type of entity in question). The latter naturally raises "meta-ascriptional" questions, concerned with the evaluation of ascription mechanisms, and Narayanan ends by suggesting a new interpretation of the imitation game at this third level.

Herbert Simon does not explicitly discuss the imitation game as such, but he approaches the question of whether machines can think in very much the same spirit as Turing, and is equally robust in the treatment of "romantic" objections. His argument is two-pronged: on the one hand, he describes a considerable body of evidence, much of it garnered from his own research, that indicates the ways in which human cognitive processing actually operates; on the other, he points to a number of computer programs, again some of them his own, that have succeeded in operating "intelligently" in strikingly similar ways. He concludes that "we need not talk about computers thinking in the future tense; they have been thinking (in smaller or bigger ways) for 35 years". Simon's work indicates that the two interpretations of Turing distinguished by Whitby may not be so far apart - here we have investigations into the nature of human intelligence providing essential theoretical groundwork for the achievement of cognitive performances by computers that amount to much more than mere imitation.

A Theoretical Barrier to Computer Thought?

Though a champion of machine intelligence, Turing was of course one of those responsible (along with Kurt Gödel and Alonzo Church) for turning the world of logic upside down during the 1930's with a series of crucial negative results regarding the theoretical powers of computers and logical systems. In 1961, this irony was exploited by John Lucas in a famous - some would say notorious - paper entitled "Minds, Machines and Gödel" (in the journal *Philosophy*), where he argued that the essential limits on formal computation implied by such results provided solid proof that human thought, which successfully discovered these limits, cannot be reduced to algorithmic processing, and hence cannot be copied (in at least this crucial respect) by a computer program. Lucas' article provoked a host of replies and "refutations", and also inspired a number of other notable thinkers (most

recently the theoretical physicist Roger Penrose) to pursue a similar line. For this volume he has written a typically forthright and uncompromising "Retrospect" on the debate, in which he aims to refute his critics and confirm what must be, if successful, the most striking and fundamental conclusion about human nature ever to be drawn from a result in mathematical logic.

Responding to Lucas and Penrose on Turing's behalf, Robin Gandy considers some examples of what might be supposed to be non-algothmic "divine spark" thinking in mathematics and logic - for example the proof of Gödel's second theorem. He provides a fascinating insight into what goes on in the mathematician's mind when attempting to grapple with abstract objects such as infinite sequences, but dismisses the idea that the "spark", when it comes, is really "divine" or in any other way essentially resistant to mechanical modelling. It may indeed not be strictly *algorithmic* - but that is because it is *fallible* (and hence non-effective, in the technical sense), being based on such trains of thought as "I see how it goes ...", "Wouldn't it be nice if ...", "This looks rather like that ..." and so on, rather than because it is *non-mechanical*. Gandy ends his paper by "coming down off the fence on both sides", though his vision of the future of computers (as potential mathematical colleagues for example) is strikingly reminiscent of Turing's.

Turing Computability and the Church-Turing Thesis

Speculation regarding the existence and potential fruitfulness of non-algorithmic thinking naturally raises issues concerning the scope and limits of what can properly be called algorithmic. And here we come to what is arguably Turing's crowning achievement - the definition of a precise notion of computability in terms of Turing Machines, and the application of this to prove the unsolvability of Hilbert's *Entscheidungsproblem*. Here we are not concerned so much with the technical details of Turing's work but with its philosophical implications, and in particular with the question whether he is right to claim that our "intuitive" concept of effective computability is completely exhausted by Turing Machine computability (and the other precise notions of computability that have been proved to be equivalent, such as general recursiveness and λ-definability)? This is the famous Church-Turing thesis, and provides the topic of Antony Galton's paper.

Galton explores the difficulties surrounding the interpretation and evaluation of the Church-Turing thesis (CT), some of which spring from its apparent oddity in asserting an equivalence between

a precise notion (Turing Machine computability) and a vague one (the idea of a computation which is intuitively "effective"). This raises questions regarding its status: should CT be seen as a conceptual claim, or an empirical assertion, or even instead as a stipulation - a proposal to replace the vague intuitive notion with the precise one? To shed light on these issues Galton adopts an interesting approach based on the semi-precise concept of *black-box computability*, which is intended to provide a more tractable substitute for the inituitive notion, whilst retaining an appropriate level of indeterminacy. After considering its implications for CT, and a variety of possible conceptual revisions prompted by recent work (e.g. of David Deutsch and Chris Fields on Quantum Computability, and Iain Stewart's paper in this volume), Galton ends by discussing the relevance of CT for research in Artificial Intelligence.

Chris Fields, one of those whose earlier work is mentioned by Galton, here addresses the issue of what it is for a process to count as a computation, an issue which clearly has major significance for the interpretation and assessment of the Church-Turing thesis. Fields suggests an intimate connection between computation and measurement: a physical system can be considered as behaving computationally only when its states are measured, and the measurements are interpreted, in particular ways. This has an important implication, because the methods of measurement and interpretation employed by Computer Science are independent of whether the system under study is natural or an artifact. Fields concludes that whether a system is to count as a computer is a pragmatic question, to be answered by considering the explanatory utility of computational descriptions of its behaviour.

Aaron Sloman covers similar ground to Fields, drawing some similar conclusions. He is concerned to make clear what counts as a "computational process", where such processes are to be understood as what underlies intelligence. In this sense, he insists, the notion is broader than Turing Machine computability, but unfortunately there is no clear way to delineate it, since if our definition is extended to include all the kinds of processes that play a role in intelligence, it will become hard if not impossible to draw any line between computational and non-computational processes without falling into circularity. He therefore recommends abandoning the idea that any precisely defined concept of computation can be the key notion underlying intelligence, and instead recommends the study of a variety of architectures and mechanisms, with the aim of developing a new theory-based taxonomy of cognitive processes.

Beyond the Turing Machine: New Horizons

The Turing Machine provided the first clear, precise and determinate specification of a computing machine, and has since proved itself to be an immensely valuable reference point for developments in computability and complexity theory. But Iain Stewart argues that in some important areas of the latter field, at least, the Turing Machine has had its day, and could profitably be replaced as an analytical tool by an appropriate formal logic. This makes the representation of a problem more natural, and also its transformation into an executable high-level program. It also has potential pedagogic advantages, in demonstrating how the study of complexity theory has genuine relevance to practical computer programs rather than just to the apparently artificial workings of a theoretical Turing Machine. Stewart readily concedes, however, that Turing's brainchild will retain its place as a more general unifying concept, given the inability of his own logical treatment to deal with problems of arbitrary complexity.

Peter Mott's paper is in somewhat the same spirit as Stewart's, in that it advocates a move from an artificial to a more natural medium of representation for the treatment of complex problems. Here, however, the field is commonsense reasoning rather than complexity theory, and formal logic now plays the role of target rather than proposed replacement. The traditional Montague paradigm for modelling linguistic reasoning, which Mott opposes, involves the translation of sentences into logical formulae, with inferential operations being performed on those formulae, and conclusions finally being retranslated back into sentences. Mott recommends instead the direct use of natural language as an inferential medium, cutting out the logical middleman by means of what he calls "Grammar Based Inference". This can dramatically reduce the complexity of commonsense reasoning, though admittedly at the cost of some loss of rigour, and it no doubt provides a far more plausible model of how our own reasoning actually takes place. It is also tantalisingly reminiscent of ancient syllogistic logics - raising the intriguing possibility that such logics, rejected in the past because of the alleged complexity and multiplicity of their forms, could ultimately be rehabilitated in the computer age by the recognition that mere linear complexity of forms is quite insignificant compared with the intractable exponential complexity that can result when we try to model reasoning using our traditional formal systems.

Joseph Ford is well known as the "Evangelist of Chaos", spreading the word of this new scientific "paradigm" with an enthusiasm and style characteristic of the Southern states. His paper provides an engaging introduction to this exciting field, with useful pointers for those who wish to pursue its potentially dramatic implications for the theories of complexity, computability and information. Ford particularly draws attention to the important discovery by Greg Chaitin of an

information theoretical analogue to Gödel's Theorem, proving "that there exist naturally occuring, simple questions whose answers are so complex they contain more information than exists in all our human logical systems combined". Physics provides such questions, and Ford examines some of the uncomfortable implications of the new paradigm for physical measurement before suggesting, with the aid of two examples, that chaos may nevertheless provide an ally rather than an enemy in the attempt to tame complexity, enabling us "to solve incredibly complex problems by letting controlled chaos do the work". He then provides a brief discussion of the place of chaos in the quantum world (which according to Penrose provides the essential substructure of human intelligence) before concluding with a typically evangelistic coda.

The collection ends with yet another example of work illustrating the great breadth of philosophical ideas taking significant inspiration from Turing. This is the development, pursued by Clark Glymour amongst others, of a mathematical "theory of discovery". The extent of Turing's influence here is perhaps surprising, because the famous negative results which he, Gödel and Church established, by ruling out an algorithmic decision procedure even for the limited domain of first order logic, may have seemed to close the door on any interesting formal treatment of epistemology. But Glymour sketches how the notion of "knowledge in the limit", formulated by Gold and Putnam in 1965, has provided the key to open a new and fascinating field of investigation, which builds on the theories of mathematical logic, computation, and recursion, and has potential implications in many areas including not only the Philosophy of Science, but also Artificial Intelligence, Cognitive Neuropsychology, Economics and even, intriguingly, the formal treatment of epistemological Relativism.
