Andrew Hodges[*]

Did Church and Turing Have a Thesis about Machines?

This article draws attention to a central dispute in the interpretation of Church's Thesis. More precisely, we are concerned with the Church-Turing thesis, as it emerged in 1936 when Church endorsed Turing's characterization of the concept of effective calculability. (The article by Sieg in this volume details this history. It is valuable also to note from Krajewski, also in this volume, that the word 'thesis' was used only in 1952.) This controversy has a *scientific* aspect, concerning the nature of the physical world and what can be done with it. It has a *historical* aspect, to do with the 'confluence of ideas in 1936'. We shall focus on the historical question, but it is the continuing and serious scientific question that lends potency to the history.

The principal protagonist in this matter is the philosopher B.J. Copeland, who when writing with his colleague D. Proudfoot for a wide readership in *Scientific American*, denounced a prevailing view of Church's Thesis as 'a myth' [Copeland and Proudfoot 1999]. Copeland has made similar assertions in numerous leading articles for journals and works of reference, e.g. [Copeland 2000; 2002; 2004]. What is this myth? It is that the Church-Turing thesis places any limitation on what a machine can do. On the contrary, according to Copeland and Proudfoot, 'Church and Turing claimed only that a universal Turing machine can match the behavior of any human mathematician working with paper and pencil in accordance with an algorithmic method—a considerably weaker claim that certainly does not rule out the possibility of hypermachines.' 'Hypermachines' are defined by Copeland to be physical machines that outdo Turing

[*]A. Hodges, Wadham College, University of Oxford, Oxford OX1 3PN, U.K., <andrew.hodges@wadh.ox.ac.uk>.

computability; Copeland and Proudfoot insist that Turing 'conceived of such devices' in 1938.

The origin of this argument lies in the work of the logician Robin Gandy, who had himself been Turing's student. His article 'Principles of Mechanisms' [Gandy 1980] distinguished Church's thesis from what he called 'Thesis M', the thesis that what a machine can do is computable. Gandy [1988] also emphasised that Turing's original argument was drawn from modelling the action of a human being working to a rule, and was *not* based on modelling machines, as may indeed readily be seen from [Turing 1936]. Gandy criticised Newman [1955] for saying that Turing embarked on 'analysing the general notion of a computing machine'.

There are good reasons for Gandy's emphasis on this distinction, and for others to follow him in emphasising the model of human computation. One is that by the 1970s it was quite a common assumption that the digital computer already existed when Turing made his definition, and that he had written down an abstract version of it. This grossly understates Turing's achievement: the digital computer arose only ten years after Turing wrote his paper, and it can be argued that his 'universal machine' supplied the principle on which the digital computer was based (directly, in his own plans, and indirectly, in von Neumann's.) Another reason is that it must be appreciated that Turing was addressing Hilbert's question about methods that can be applied by human mathematicians. Another reason lies in observing that Turing's discussion of human memory and 'states of mind' makes his 1936 work basic to the cognitive sciences, and in particular to his own later discussion of artificial intelligence.

Despite all this, however, this distinction was *not* clearly drawn by Church or Turing in that period of the 'confluence'. The evidence comes from Church's review of [Turing 1936] in which he endorsed Turing's definition. Indeed the main point of this article is simply to bring this review [Church 1937a] to greater prominence. A full transcription of the review is given in [Sieg 1997, in this volume]. We need only the relevant opening paragraph.

> The author [Turing] proposes as a criterion that an infinite sequence of digits 0 and 1 be "computable" that it shall be possible to devise a computing machine, occupying a finite space and with working parts of finite size, which will write down the sequence to any desired number of terms if allowed

> to run for a sufficiently long time. As a matter of convenience, certain further restrictions are imposed in the character of the machine, but these are of such a nature as obviously to cause no loss of generality—in particular, a human calculator, provided with pencil and paper and explicit instructions, can be regarded as a kind of Turing machine.

It is apparent that Church was unaware of Gandy's distinction between the Church-Turing thesis and Thesis M. Indeed, if Church had actively set out to cultivate the 'myth' strenuously denounced by Copeland, he could hardly have done so more effectively. For Church's words not only referred to machines, but actually claimed a *definition* of computability in terms of the properties of machines, considered as 'devised' objects with a 'size' in 'space'. Note that Church could not have been using the word 'machine' with an implicit restriction to the Turing machine, because, by definition, he was introducing this new concept to readers ignorant of it. (Indeed, the expression 'Turing machine' was coined in this review.) 'Computing machine' here means any machine at all (of 'finite size') which serves to calculate. This assumed generality is confirmed in the immediately following article [Church 1937b] in the *Journal of Symbolic Logic,* where Church reviewed Post's independently conceived formalism of a rule-following 'worker'. Church criticised Post for requiring a 'working hypothesis' that it can be identified with effective calculability. He specifically contrasted Post's formalism with Turing's, and referring to Turing's paper, wrote:

> To define effectiveness as computability by an arbitrary machine, subject to restrictions of finiteness, would seem an adequate representation of the ordinary notion, and if this is done the need for a working hypothesis disappears.

It was therefore the very *generality* of Turing's machine concept, not its particular formalization, that led Church to commend it.

Copeland goes much further than Gandy and holds that Church and Turing positively *excluded* machines from their thesis. How can Copeland reconcile his assertion with the fact that Church based his observations on the concept of 'an arbitrary machine'? In [Copeland 2002] there is no mention of Church's review of [Turing 1936], and the problem is thus avoided. But Copeland does there cite the immediately following review of Post, with the same quotation as given above, and with the following gloss:

> [...] he is to be understood not as entertaining some form of thesis M but as endorsing the identification of the effectively calculable functions with those functions that can be calculated by an arbitrary machine whose principles of operation are such as to mimic the actions of a human computer. (There is much that is 'arbitrary' about the machines described (independently, in the same year) by Turing and Post, for example the one-dimensional arrangement of the squares of the tape (or in Post's case, of the 'boxes'), the absence of a system of addresses for squares of the tape, the choice between a two-way and a one-way infinite tape, and, in Post's case, the restriction that a square admit of only two possible conditions, blank or marked by a single vertical stroke.)

This gloss, with its bizarre interpretation of the word 'arbitrary', achieves Copeland's reconciliation only by reversing the sense of Church's statement. Church specifically described Turing's human calculator as a particular example of a machine, not as the definitive form. Note that Church's explicit use of the word 'human' confirms that his general setting for effective calculation is *not* necessarily human.

As a summary of [Turing 1936], Church's review was notably *incorrect*. Turing had not even referred to machines of finite size, let alone defined computability in terms of their alleged powers. One might criticise Church's review in other ways: he omitted the emulation of mental states which is a striking feature of Turing's analysis. Stipulating a finite *number* of working parts, rather than a finite size, would better indicate the finite number of configurations of a Turing machine. Moreover, Church's over-briefly stated condition of finiteness fails to bring out that the working space (the 'tape') must not be limited. Only a finite amount of space is used in any calculation, but there is no preset bound on how much may be demanded. A completely finite machine must repeat itself: one aspect of Turing's breakthrough was that he saw how to keep a finiteness of specification but escape this limitation.

I owe to Wilfried Sieg [2005] the observation that Gödel followed Church and ascribed to Turing an analysis of machines. There seems to be no obvious answer to the question of why both Church and Gödel imputed to Turing's analysis something that was not actually there. Another question arises when we imagine Turing at the Graduate College in Princeton in 1937, reading Church's review. Did he recoil with horror, seeing it as a travesty of his achievement, or did he see it as a legitimate variant or development of his theory? If Turing *had* regarded it as seriously misrepresenting his ideas, he would in not have been deterred from saying so by Church's seniority. He was shy socially but very confident of his own judgment in all sorts of matters. (Thus, regarding another kind of Church, he wrote in 1936: 'As for the Archbishop of Canterbury, I consider his behaviour disgraceful'.) But he recorded no dissent.

If Turing had wished politely and properly to distance himself from Church's version of his definition, and re-assert his own, he had the opportunity in his 1938 doctoral thesis, subsequently published as [Turing 1939]. Yet in that paper, when giving his own statement of the Church-Turing thesis, he simply characterized a computable function as one whose values can be found 'by some purely mechanical process', saying that this may be interpreted as 'one which could be carried out by a machine'. This does not have the full force of Church's words 'arbitrary machine', for the words 'a machine' could be read as meaning 'a Turing machine', but it notably makes no effort whatever to alert the reader to any distinction between 'machine' and 'human rule-follower.' It is hard to see how Turing could have left his wording in these terms if he had regarded Church's formulation as a serious and misleading error. Morever, Church also simply repeated his 'machine' characterization of computability in a later paper [Church 1940], which does not suggest that Turing had ever expressed an objection to it while they were in contact at Princeton.

It appears that Church and Turing (and others, like Gödel and Newman) used the word 'machine' quite freely as a synonym for 'mechanical process', without clearly distinguishing the model of a mechanical process given by the human rule-follower. In fact, Church's review did not offer an *absurd* distortion or extrapolation of what Turing had done. With some sketch of what was assumed about 'machines' and what was meant to be 'obvious' about the complete generality lying behind Turing's formalization, it could have been justifiable. The work of Gandy [1980] showed that under quite reasonable conditions on what is meant by 'machine', his Thesis M is actually true. Sieg [2002] has extended and improved upon Gandy's results. The main point is that their conditions allow for machines which are not restricted to making one step at a time, but perform parallel computations.

Even so, Gandy and Sieg's analyses are far from being an exhaustive account of what a physical machine might be. They do not allow for the phenomenon of entangled quantum states, which is already of technological importance in quantum cryptography. For this reason alone, this type of logical analysis lags behind modern physics. Computer science depends upon the implementation of the logical in the physical, and the review of the distinguished computer scientist A.C.-C. Yao [2003] shows the depth and range of physical process now seen as relevant to its future progress. It is worth noting that Yao defines the Church-Turing thesis in terms of what can be computed by 'any conceivable hardware system', saying that 'this may not have been the belief of Church and Turing, but it has become the common interpretation' of their thesis. Yao regards the thesis not as a dogma but as a claim about physical laws which may or may not be true. Yao's careful words about what Church and Turing believed are fair: we cannot know quite what they thought, but the evidence points to a standpoint closer in spirit to Yao's than to Copeland's. Yao is not quite so careful, however, in his statement of the thesis, for he omits to include a 'finiteness' condition such as Church emphasised. Some such condition is obviously essential—an infinitely long register of data must, for instance, be ruled out.

Turing's silence on the question of 'arbitrary machines' is rather surprising because he was in many ways an outsider to the rather isolated logicians' world, having a broad grounding in applied mathematics and an interest in actual engineering. On the specific question of restriction to serial working, it is noteworthy that he had already discussed in [Turing 1936] how human mental 'scanning' of many symbols at once could be reduced to a serial process. Thus he could very well have initiated the kind of theory of machines later undertaken by Gandy—the more so since machines with parallel action ('simultaneous scanning', he called it) were crucial to Turing's success with the Enigma in 1939–40. On broader issues too, he was well-qualified to point out that in 1937 formulations such as 'parts' of a machine and 'sufficiently long time' were already obsolete and demanded much more serious analysis: twentieth-century physics had transformed the classical picture of space, time and matter which Church's words appealed to. At the age of sixteen he had understood the basis of quantum entanglement and of curved space-time, and there was nothing to prevent him drawing attention to the questions thereby aroused. (Curiously enough, Gödel later found a solution of Einstein's equations which exhibits closed timelike lines, a fact which in itself shows that the concept of 'sufficiently long time' is unsatisfactory without more refined analysis.)

Turing's background in physics did in fact re-assert itself later on. First, in his individualistic trajectory, came his own engineering of machines at Princeton; then came an extensive wartime experience of electromagnetic and electronic machines which led to his digital computer design in 1945. In 1948, Turing's report 'Intelligent Machinery' gave a brief analysis of 'machines' which did take note of a necessary grounding in physical concepts (for instance thermodynamics and the speed of light). In this paper Turing simply summarised computability using the phrases 'rule of thumb' and 'purely mechanical' as equivalents, without drawing a distinction between the human rule-follower model and the machine model. Indeed, Turing drew these ideas together in a discussion of 'Man as a Machine' and the brain as a physical system. In his edition of Turing's papers, Copeland [2004, p. 480] acknowledges that Turing wrote that a computer could replace 'any calculating machine', but explains this by saying that Turing 'would' have characterized a calculating machine as doing only what could be done by a human computer. But Turing never actually gave this definition, and indeed Turing [1948] gave his readers the reverse image: he described a program, to be worked out by a human rule-follower, as a 'paper machine'.

However, in this 1948 paper, and thereafter, Turing did refine the concept of 'machine'. He distinguished 'active' machinery from 'controlling' machinery, giving 'bulldozer' as an example of the former; the latter type, which we would probably now call 'information-theoretic', is the subject of his discussion. (Thus, we are concerned with abstracting what it is that makes a machine 'mechanical', not with its physical action.) Turing also distinguished 'continuous' from 'discrete' machines, and again it is the latter with which we are principally concerned. Turing's main argument, both in this 1948 paper and in his very famous publication [Turing 1950], was that the action of the brain can be captured by a discrete 'controlling' machine. Of course, Turing by now had a more developed theory in which more 'intelligent' machine behaviour would be acquired through dynamical interaction with the environment, but it was all still within the arena of the discrete machine and governed by computable operations. Turing [1948] was at pains to point out that the brain is actually continuous, even though there is 'every reason to suppose' that a discrete model will suffice to model it. In [Turing 1950] he gave a more explicit argument for this supposition. Thus Turing began to raise questions about connection of the computable and discrete with continuous physics.

Since the 1950s many leading figures (e.g. Weyl, Kreisel, Wigner, Feynman, Chaitin) have raised questions about the physical basis of the Church-Turing thesis. Many articles in this volume indicate the range of ideas now studied. One notable contributor to this broader picture is Roger Penrose, who argues that there must be an uncomputable aspect to the physics of quantum measurement [Penrose 1989; 1994]. Interestingly, Turing [1951] showed evidence of contemplating just this possibility. In this radio talk, mainly rehearsing his earlier arguments about modelling the brain by a Turing machine, Turing inserted a new point that the uncertainty in quantum mechanics might make this impossible. This single sentence, which stands in contrast to his 1950 assertions, is the only actual reference by Turing to an aspect of physical law that might be uncomputable. In his last year of life, Turing also started an investigation of the quantum measurement process [Gandy 1954]. Thus if we look at a longer time-scale, we can see Turing as helping to open the whole question of computability and physics as it has slowly developed over the last 50 years, a point developed in [Hodges 2004]. However, Copeland's contention is specifically concerned with the meaning of what was formulated in 1936. He holds both that the Church-Turing thesis is true, and that physical machines may be capable of computing uncomputable functions. The only way to reconcile these statements is to assert that Church and Turing positively held in 1936 that their concept of effective calculation did not refer to machines. The historical record does not support this contention.

As we noted at the outset, there are both historical and scientific questions involved in this issue. One cannot separate them entirely because the views of great founding figures are of special significance and deserve to be studied. The importance of originators is reflected in the way Copeland enlists Turing in the cause of hypothesising machines which might perform uncomputable tasks, writing of Turing's allegedly ‘forgotten ideas.’ Copeland and Proudfoot specifically assert that Turing’s ‘oracle-machine’ [Turing 1939] is to be regarded as such a machine, suggesting various ways in which it could be physically realised, e.g. as a quantity of electricity to be measured with infinite precision [Copeland and Proudfoot 1999]. Copeland and Proudfoot do not explain how this infinitude could possibly be effected in accordance with any known physical principle, and of course there is no suggestion of any such thing in [Turing 1939]. They nevertheless announce this implemented oracle-machine as a potential technological revolution as great as that of the digital computer, crediting the ‘real Turing’ with this vision [Copeland and Proudfoot 1998]. They see the prevailing ‘myth’ about the Church-Turing thesis as an impediment to realising this ambition. These assertions, scientific and historical, are alike ill-founded.

References

Church, A. [1937a], Review of Turing [1936], *Journal of Symbolic Logic* **2**, 42.

Church, A. [1937b], Review of Post [1936], *Journal of Symbolic Logic* **2**, 43.

Church, A. [1940], “On the Concept of a Random Sequence”, *Bull. Amer. Math. Soc.* **46**, 130–5.

Copeland, B.J. and Proudfoot, D. [1998], *Enigma Variations*, London: Times Literary Supplement, 3 July 1998.

Copeland, B.J. and Proudfoot, D. [1999], “Alan Turing’s Forgotten Ideas in Computer Science”, *Scientific American* **253:4**, 98–103.

Copeland, B.J. [2000], “Narrow Versus Wide Mechanism: Including a Re-Examination of Turing’s Views on the Mind-Machine Issue”, *Journal of Philosophy* **96**, 5–32.

Copeland, B.J. [2002], *The Church-Turing Thesis*, in *Stanford Encyclopedia of Philosophy*, (E.N. Zalta ed.), <http://plato.stanford.edu/>.

Copeland, B.J. (ed.), and Turing, A.M. [2004] *The Essential Turing* Oxford: Oxford University Press.

Gandy, R.O. [1954], letter to M.H.A. Newman, in the Turing Archive, King’s College, Cambridge, included in the *Mathematical Logic*, vol. *The Collected Works of A.M. Turing*, (R.O. Gandy and C.E.M. Yates eds.), Amsterdam: North-Holland, 2001.

Gandy, R.O. [1980], "Principles of Mechanisms", in *The Kleene Symposium*, (J. Barwise, H.J. Keisler, and K. Kunen eds.), Amsterdam: North-Holland.

Gandy, R.O. [1988], "The Confluence of Ideas in 1936", in: *The Universal Turing Machine: a Half-Century Survey*, (R. Herken ed.), Berlin: Kammerer & Unverzagt.

Hodges, A. [2004], "What would Alan Turing have done after 1954", in *Alan Turing: Life and Legacy of a Great Thinker*, (C. Teuscher ed.), Berlin: Springer.

Newman, M.H.A. [1955], "Alan Mathison Turing", *Biographical Memoirs of the Fellows of the Royal Society* **1**, 253–263.

Penrose, R. [1989], *The Emperor's New Mind*, Oxford: Oxford University Press.

Penrose, R. [1994], *Shadows of the Mind*, Oxford: Oxford University Press.

Sieg, W. [1997], "Step by Recursive Step: Church's Analysis of Effective Calculability", *Bulletin of Symbolic Logic* **3**, 154–180; also in this volume.

Sieg, W. [2002], "Calculations by Man and Machine: Conceptual Analysis", in *Reflections on the Foundations of Mathematics: Essays in Honor of Solomon Feferman*, (W. Sieg, R. Sommer, and C. Talcott eds.), *Lecture Notes in Logic* **15**, pp. 390–409.

Sieg, W. [2005], "Gödel on Computability", to appear in *Philosophia Mathematica*.

Turing, A.M. [1936], "On Computable Numbers, with an Application to the *Entscheidungsproblem*", Proc. Lond. Math. Soc. (2), 42, pp. 230–265.

Turing, A.M. [1939], "Systems of Logic Based on Ordinals", Proc. Lond. Math. Soc. (2), 45, pp. 161–228.

Turing, A.M. [1948], "Intelligent Machinery". Report for the National Physical Laboratory, published in various forms since 1968, e.g. in [Copeland and Turing 2004].

Turing, A.M. [1950], "Computing Machinery and Intelligence", *Mind* **49**, 433–460.

Turing, A.M. [1951], “Can Digital Computers think”, BBC radio talk, available in [Copeland and Turing 2004].

Yao, A.C.-C. [2003], “Classical Physics and the Church-Turing Thesis”, *Journal of the ACM* **50**, 100–105.

Added note: Since this article was written, Professor Copeland has advanced the discussion in this volume. He has widened the scope, and I thank the Editors for permitting a comment on one of his central additional arguments. This (page 162) rests on Turing's 1950 discussion of computers. The ‘unlimited store’ described by Turing does not correspond, as Copeland asserts, to ‘an unlimited number of configurations’ in a Turing machine table of behaviour. This is because Turing's 1950 explanation does not present a digital computer's storage as analogous to the tape of a Turing machine. Instead, Turing *omits the tape*, and presents all the storage as internal to the machine. This makes it difficult to explain the full scope of computability, which requires the concept of unlimited tape. He has to refer to an ‘unlimited store’ instead. So his ‘unlimited store’ corresponds to the unlimited *tape* of a standard Turing machine, not to its configurations. (Turing says of the unlimited store that ‘only a finite amount can have been used at any one time’, just as with the storage on a Turing machine tape.) What Turing in 1950 calls a theoretical ‘infinitive capacity computer’ is the (universal) Turing machine of 1936. Its ‘states’ include what in the standard Turing machine description are states of the *tape*, which are indeed generally unbounded. Those discrete state machines with only a *finite* number of possible ‘states’—the condition that Copeland italicises as vital evidence—correspond to Turing machines which use only a finite tape, or equivalently, to totally finite machines which need no tape at all. Nothing here goes beyond computability. Rather, it emphasises the *finite* resources that Turing was discussing as necessary for mental behaviour—giving a figure of not much more than 10^{10} bits of storage.

Index