

Doing Justice to the Imitation Game ; a Farewell to Formalism

By Jean Lassègue

Abstract

My claim in this article is that the 1950 paper in which Turing describes the world-famous set-up of the Imitation Game is much richer and intriguing than the formalist ersatz coined in the early '70s under the name "Turing Test". Therefore, doing justice to the Imitation Game implies showing first that the formalist interpretation misses some crucial points in Turing's line of thought and secondly that the 1950 paper should not be understood as the *Magna Charta* of strong Artificial Intelligence but as a *work in progress* focused on the notion of Form. This has unexpected consequences about the status of Mind, and from a more general point of view, about the way we interpret the notions of Science and Language.

*

The common use of the phrase 'Turing test' instead of the expression 'Imitation game', preferably used by Turing [1], is concomitant with a shift of meaning in the expression 'Artificial Intelligence'. During and after World War II, the latter expression would rather mean the mechanical gathering of enemy information. But the meaning began to evolve at the end of the '50s when 'Artificial Intelligence' was used to raise the standard of revolt against 'Cybernetics' [2]. Contrary to Cybernetics which had always considered the physical and the logical level of cognition at the same time [3], AI was based upon a radical cut between a purely formal, computable level of intelligibility and its physical counterpart which could be passed over and left to engineering. Later, in the mid '70s, the notion of a 'Turing test' became popular when mainstream Cognitive Science was claimed to be entirely based upon the brain-computer paradigm and its distinction between hardware and software derived from AI - hence a global shift towards formalism and functionalism. Interpreted from a formalist point of view, the Imitation game set up by Turing in his 1950 world-famous article amounts therefore to a test that would show, on a statistical basis, that the discrimination between the verbal expressions of a human being and those of a computer lies beyond human decision. The major claim of my paper is that this test is not feasible in the conditions just mentioned, contrary to what an abundant literature [4], blinded by a formalist interpretation, has been eager to show. This is why I consider the reduction of the imitation game to a test as an impoverishment of Turing's original point of view. To my mind, the phrase 'Turing test' is somewhat of a misnomer and I shall use three anti-formalist arguments against its use. Let me begin first by describing the origins of the formalist point of view I want to re-examine.

1. Clockwise determinism : the formalist interpretation of the Imitation Game

Let us start with the historical rise of the formalist point of view in scientific knowledge.

11. Determinism in physics and in mathematics : the Laplacian worldview

Formalism is not the ultimate paradigm of scientific rigour it is sometimes claimed to be. It has to do with a specific state of scientific knowledge which was current between the 17th and the 20th century and in which the clarification of what could be rigorously *predicted* was linked to the idea of *computation* [5]. In order to make this clear, I shall amply rest upon the epistemological framework developed by two professional logicians, G. Longo [6] and J.-Y. Girard [7], in which a parallel is drawn between the evolution of determinism in physics and that of computation in mathematics [8]. Let us begin by quoting G. Longo on Laplace and Hilbert [9] :

"Laplace proposed a paradigm for the mathematical analysis of Physics, the so-called 'Laplacian determinism'. In this perspective, the systems of (differential) equations could 'completely' describe the physical world. More precisely, if one wanted to know the state of the physical world in a future moment, with a given approximation, then it could suffice to know the current state of affairs up to an approximation of a *comparable order of magnitude*. By formally computing a solution of the intended equations, or by suitable approximations by Fourier series (as it will be said later), one could deduce (or predict or decide) the future state, up to the expected level of approximation. [...] About one century later, D. Hilbert resumed Laplace's program in a different context. He first set the basis for the rigorous notion of "formal system", as well as for the distinction between 'theory' and 'metatheory'. He then conjectured that the key system for Number Theory, Peano's Arithmetic (where he had interpreted Geometry, 1899), was complete with regard to the intended structure of numbers (or that any well formed proposition about the "world of numbers" could be decided by formal or 'potentially mechanisable' tools)."

The common picture that emerges from these two scientific views is that it is possible, from a system of equations or a system of axioms, to *decide* whether a physical or a mathematical event would occur or not : the *temporal* evolution of a dynamical system or the *logical* consequences of a formal system receive *the same deterministic* treatment. At the centre of this scientific picture we find the notion of *function* which is essential to the constitution of mathematical physics. The notion of a function was still akin in the 17th and 18th century to the idea of a step by step computation procedure [10]. Completeness and decision in systems of equations or of axioms seemed, from this point of view, to go hand in hand : *all* the events and *every* one of them are potentially computable. In the natural philosophy of that time, deterministic computation was heralded the most general model applicable to both Mind and Matter. It had a technological counterpart in the notion of a *clock* : the entire universe was nothing but a gigantic clock and the mechanics of Mind were just a very complex part in it [11]. Even if the notion of a function had tremendously evolved in between, Turing inherited the physical and mathematical worldview developed during this classical period and made a great contribution to its mathematical aspect by determining what was meant by 'potentially mechanisable'. But he also participated in transforming the key concept of determinism when he managed to show that it had an *inner limit* in the axiomatic domain. Turing was not the only one to contribute to this transformation (Gödel is of course the great figure in this case). It must be stressed that they were, so to speak, in a position where they were able to grasp both sides of the evolution of the concept of determinism - absolute determinism on the one side and relative determinism on the other. Secondly, the transformation of this concept happened in mathematics as well in physics. In physics too, Poincaré and the Andronov's school in Gorki had guessed and then established that absolute determinism was not the last chapter in physics.

These two points which are closely related to one another [12] are to be remembered when the Imitation Game is approached.

12. Formalism in mathematics and logic : a trick of the '20s

What is strictly defined as 'Formalism' in mathematics is only a technical strategy used by some mathematicians in the first decades of the 20th century in order to bypass some very deep problems encountered in the interpretation of geometry. This point has been described at length by historians of mathematical logic [13] and I shall only summarize the different stages of the story.

In the beginning was the very deep crisis triggered by non-Euclidean geometries at the end of the 19th century : the notion of space became so confused that it lost its precise meaning and the entire mathematical world was immersed in a kind of geometrical "delirium", as Frege put it [14]. Geometry, which had been the very foundation of mathematical certainty since its Euclidean axiomatization, had lost its cardinal virtue : the certainty of Euclidean geometry was not the touchstone of objective truth any longer since contradictory propositions were possible between the different axiomatics of geometry. Even geometry could be inconsistent - but if geometrical certainty was shattered for ever, where was mathematical consistency to be found again ? *Formalism* was a way out of this crisis : it was both a linguistic and an arithmetical turn in which formal coding based on natural number arithmetic became the key word. Peano and Padoa began the great linguistic reform in 1889 by reinforcing mathematical rigour concerning the basic notions (definitions, axioms, etc) used in the axiomatic theory of number [15]. Interpreting the axioms of geometry in a completely revolutionary way [16], Hilbert was able to reduce them to arithmetic by 1917 [17], delimiting the problem of consistency to that of arithmetic. In order to secure mathematics as a whole and more specifically arithmetic, he opposed meaningful propositions in contentual axiomatics to meaningless strings of symbols in formal axiomatics (called 'meta-mathematics') as content to form. It became possible to manipulate strings of symbols without taking care of their supposed meaning, *even in arithmetic*. A mathematical core was therefore secured which was limited to finitary strings of meaningless symbols : even propositions in which the meaningful content used the notion of second-order infinity (i.e. richer than the indefinite enumeration of natural numbers) and which were very common in geometry, were nonetheless considered, from a formal point of view, as finite strings of symbols and could therefore be handled, at least theoretically, in the same way as propositions endowed with a finitary content. A purely formal approach to axiomatics was launched : formal axiomatics became a kind of machinery marshalling theorems generated by explicit finitary rules and would then replace geometry as the universal dispenser of truth. In 1928, Hilbert was able to formulate in a precise way some pending questions concerning determinism in formal systems [18] : it was the acme of what was to be called "Hilbert's program". What is really impressive with Hilbert's program is that it is so precisely stated that it can be *proved* right or wrong - depending on the opened questions under consideration. In a way, Turing's imitation game inherited this "informal rigour" but this does not imply that we should consider it only as a by-product of formalism. Hilbert's new approach to mathematics had far-reaching consequences concerning the relationship between mathematics and physics on the one hand and mathematics and logic on the other. I shall mention only two of them.

12.1. A consequence in the philosophy of physics

A theoretical split began to appear between foundational research in mathematics and in physics. Since Cartesian times, the intuition of the naturalness of the three-dimensional space had

been baffled by new advances in algebra. Although it was still controversial at the time, the Cantorian set-theoretical point of view, even though still controversial [19], made possible the reduction of any geometrical figure to sets of points : any dimension could then be mapped onto the straight line [20]. But the concept of a set of points had a very serious drawback : any information concerning the *structure* of space, i.e. the *geometrical forms*, was entirely lost in the mapping process. This was pointed out by physics-oriented mathematicians like Poincaré [21] and Weyl [22] who, remembering Riemann's revolutionary way of interpreting space itself as a geometrical object [23], were not keen on encapsulating mathematics in a pure formal setting completely severed from physics. And even if Hilbert and the Hilbert school made some crucial contributions to physics, the epistemological atmosphere of the '20s developed a kind of inner schizophrenia : whereas geometry was at the forefront of research in theoretical physics, it was banned altogether in foundational mathematics. The solution to this theoretical split was reductionism [24] : the whole structure of scientific knowledge was supposed, in the end, to rest upon the finitary base delimited by formal axiomatics [25].

122. A consequence in the philosophy of logic

As for logic, the main consequence of the rejection of geometry out of foundational research was that the manipulation of meaningless strings of symbols in formal axiomatics had to rest upon an entirely new foundation. Since Space could not be trusted to the same degree any more, the notion of Mind found a new favour : as a philosophical companion to the algebraization of mathematics since Cartesian times [26], the finitary step-by-step construction, once reduced to the step-by-step construction of meaningless strings of written symbols, was then heralded the basic *modus operandi* of Mind itself. Hilbert stated [27] :

"Our thinking is finitist ; when we are thinking, a finitist process is going on."

It was a philosophical axiom, so to speak, which was made necessary by the direction axiomatic research had taken. Since the actual reduction of the entire scientific knowledge to finitary construction was not entirely achieved by the formalist program yet [28], one could say that Hilbert was, at the time, making high bids on the future foundation of Science - on the one hand, physics was deeply concerned with geometry in order to determine the ultimate properties of Matter, while on the other, mathematics, at least in its foundational enquiry, was only concerned with formal inference, i.e. the laws of Thought. The chasm between Matter and Mind that Hilbert contributed to establish accelerated the parting of the ways between the foundational paradigms accounting for the development of mathematics on the one hand and of physics on the other.

13. Turing's contributions to the formalist viewpoint : physical Determinism of the formal Mind

Turing took foundational issues concerning the inner working of the Mind where Hilbert had left them fifteen years before him, and he made two important contributions to the formalist point of view. First, he transformed the only philosophical Hilbertian axiom concerning the finitist process of thinking into a purely technical matter by inventing a formal counterpart to the notion of a mental process. Secondly, he managed to implement a purely logical concept of computation in the physical world. These two points are important because the argument devised in the Imitation Game is based upon them. Let me describe them briefly.

131. Formal modelling of mental processes

Since the Gödelian arithmetization of formal axiomatics (i.e. the one-to-one correspondence between formalized elements and natural numbers) [29], it had become clear that understanding the notion of a finitary deduction in a formal context would imply a very thorough analysis of what computation actually was. Of course, Hilbert had postulated the finiteness of the operating Mind : proving a theorem in formal axiomatics amounted for him to computing an algorithm. But the algorithmic finitary process was justified only by an external and, indeed, psychological commentary about what was meant by "carrying out a finite number of steps" in order to perform a formal deduction. What was to be formalized, i.e. what was the inner working of meaningless strings of symbols when something was computed was still unclear. Turing carried the formal point of view a step further when he managed to define with suitable precision exactly what was meant by "carrying out a finite number of steps". The mental process underlying any computation was decomposed into an explicitly finite number of steps which could therefore be completely analysed in a tabular form without any appeal to psychology [30]. In order to make this point of view clear, Turing was inviting the reader of his 1936 article to place himself in a finitary mode of thinking and to become a "computer" [31] : "The behaviour of the computer at any moment is determined by the symbols which he is observing and his 'state of mind' at that moment. We may suppose that there is a bound B to the number of symbols or squares which the computer can observe at one moment. If he wishes to observe more, he must use successive observations. We will also suppose that the number of states of mind which need to be taken into account is finite. The reason for this are of the same character as those which restrict the number of symbols. If we admitted an infinity of states of mind, some of them will be 'arbitrarily close' and will be confused. Again, the restriction is not one which seriously affects computation, since the use of more complicated states of mind can be avoided by writing more symbols on the tape."

Describing from the outside the finitist mental act would not do any more ; everyone had to carry out the experience for himself by entering a finitary and constructive state of mind. Therefore, the reader, by identifying himself with the process performed within the signs themselves, was to be convinced of the finitary nature of the source of all computation. Through this experience, Turing was reversing Hilbert's philosophical axiom : it was *the written symbols that generated states of mind* and not the other way round. Therefore, the mental act was secondary in comparison with what could be linguistically described from a finitist point of view - what was at stake was only the *mapping* of a discrete set of symbols with a set of behaviours in a computing machinery and not the 'reality' of some states of mind that were only postulated. The formal representation of this "mental act" could be carried out *no matter who or what* was actually performing it : the finitary process itself, since it was only a finite list of "behaviours" any computer could perform, was entirely mirrored in a formal treatment of written symbols, i.e. a *program*. This notion of a formal counterpart was far from the Hilbertian mentalism and was modelled according to the Gödelian arithmetical method. It was called by Turing a "machine" and soon after Church's review of Turing's 1936 article [32], a "Turing machine". One can then say that the five-year evolution (1931-1936) which goes from the Gödelian arithmetization of formal axiomatics to the Turingian building of 'tables' could be aptly labelled the birth period of programming languages. This non-mentalist and hyper-formalist attitude, so to speak, should be remembered also when we shall get to the Imitation Game.

132. Physically implemented automata

The second contribution of Turing to the formalist viewpoint was the physical implementation of the logical machine he managed to plan as early as 1944. It could be argued that physical

implementation has nothing to do with formalism *stricto sensu* but in fact, as Longo and Girard have shown, the idea of *deterministic predictability* which is common to mathematics and physics in their classical stance stands at the very center of formalism. What Turing did was to take the Laplacian worldview one step further in physically implementing a Laplacian deterministic device - the computer [33]. Therefore, the notion of an implemented universal Turing machine bridges the gap between Mind and Matter by exhibiting in a physical object *a strict determinism even Laplace would not have dreamed of*. To quote Turing himself (emphasis mine) [34] :

"It will seem that given the initial state of the machine and the input signals it is always possible to predict all future states. This is reminiscent of Laplace's view that from the complete state of the universe at one moment of time, as described by the positions and velocities of all particles, it should be possible to predict all future states. The prediction which we are considering is, however, rather nearer to practicability than that considered by Laplace. The system of the "universe as a whole" is such that quite small errors in the initial conditions can have an overwhelming effect at a later time. The displacement of a single electron by a billionth of a centimetre at one moment might make the difference between a man being killed by an avalanche a year later, or escaping. *It is an essential property of the mechanical systems which we have called 'discrete-state machines' that this phenomenon does not occur. Even when we consider the actual physical machines instead of the idealised machines, reasonably accurate knowledge of the state at one moment yields reasonably accurate knowledge any number of steps later.*"

So, in a way, the physical machine called a computer does not behave like any other natural system : following Laplacian determinism *better* than any other material object, it should be rightly called a *supernatural* object since it fits the human rationalist desire for absolute determinism. The step-by-step routine derived from the formal model of Mind casts therefore a ghostly touch on this particular physical object. This last point has, of course, some profound consequences on the formalist interpretation of the Imitation Game to which we come now.

14. The formalist interpretation of the Imitation Game

Turing's 1950 article addresses the question whether machines can be said to think [35]. In order to settle the problem experimentally and not on purely metaphysical grounds, which would try and define what intelligence *is* without *proving* what it is, Turing sets up an "imitation game" which is a kind of abstract oral examination and which will decide the matter experimentally. Although the text is extremely famous, I have thought it profitable to quote it, since its argument has been widely misrepresented [36] :

"It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A.' The interrogator is allowed to put questions to A and B . . . We now ask the question, "What will happen when a machine takes the part of A in this game ?" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman ? These questions replace our original 'Can machines think ?'"

Let me call the game played between human beings, game #1 and the game played between a woman and a computer, game #2. The Turing test is concerned with the passage from game #1

to game #2 when one substitutes the machine for the man. From the reader's point of view, the outcome of the second game sounds very formal indeed since it looks exactly like Turing's first contribution to the formalist point of view I mentioned above (§ 132) when I said that "the formal representation of the "mental act" could be carried out *no matter who or what* was actually performing it" - the physical difference between a human being and a machine disappears completely as far as the running of a program is concerned. Moreover, the reader is placed in a position which resembles an impossibility proof for a decision problem ; just as Turing proved in his 1936 article that there was no machine which could decide whether any given formula is deducible from the axioms of a formal axiomatic where arithmetic is embedded [37], here the reader is supposed to realize that the interrogator will not manage to decide whether the answers were given by a woman or a computer. The interrogator is therefore identified with the supposedly all-deciding machine Turing made use of in his 1936 proof when he showed that it would be nevertheless at a loss when it comes to decide every formula of a formal axiomatic.

So the whole Imitation Game set up looks like a formalist construction [38] : the interrogator-programmer introduces questions as *inputs* in the machine i.e. in the game ; the players play the part of *instruction tables* for this machine ; then the interrogator-programmer decides whether the answers are true or false and gives other questions as new inputs. Once this point is reached in the formalist interpretation, it is very tempting to try and draw some consequences derived from Turing's second contribution to the formalist viewpoint, namely the implementation of a Laplacian discrete-state machine in the physical world. The formalist point of view is then considered as the right level of description able to isolate in the brain what pertains to the Mind, i.e. what actually belongs to the supernatural Laplacian machinery of a computer. A. Hodges brilliantly summed up the whole formalist viewpoint from this perspective [39] :

"The Turing thesis is that the discrete-state-machine model is the relevant description of one aspect of the material world - namely the operation of brains. Turing made a robust, indeed provocative, defense of this view and its implications. Pushing his thesis as far as he could, he opened up new issues and arguments. His continuing discussion of 'thinking and 'intelligence' tended always to enlarge the scope of what was to be considered relevant. In 1936, his argument had centered on the carrying out of algorithms, in the work of 1946-1948 chess-playing (much discussed in wartime work) became his paradigm of intelligence, a principal point being that a successful chess-playing machine would have to evolve algorithms never explicitly supplied to it. In Turing 1950, the arguments turned on the eventual success of the 'intelligent machinery' in the much more ambitious task of sustaining general conversation."

Turing's project throughout his entire life seems therefore to be rightly described as the attempt to generalize the discrete-state-machine model - having started from the "ideal" level of mathematics and logic, Turing would have managed later on to apply his model to the "physical" level of the working brain. I do not agree with what I consider a formalist bias in the interpretation of Turing's research. The first reason is that it rests upon the formalist distinction between ideal Mind and physical Matter which has to be accounted for and not only presupposed. Secondly, I am not sure it fits perfectly with Turing's biography. A. Hodges has interpreted Turing's intellectual evolution as if he had gradually abandoned the problem of the uncomputable which was still under focus in his 1938 article [40], because he realized during World War II how powerful his discrete-state machine model was [41] :

"My guess is that there was a turning point in about 1941. [...] It was at this period that he abandoned the idea that moments of intuition corresponded to uncomputable operations.

Instead, he decided, the scope of the computable encompassed far more than could be captured by explicit instruction notes, and quite enough to include all that human brains did, however creative and original."

But this interpretation has, to my mind, a very serious drawback : how are we to account for the fact that Turing abandoned Computer Science altogether after 1951 - that is, just after having completed 'Computing Machinery and Intelligence' - and swapped to full-time morphogenetic research in which computer science was playing a part as a modelling tool but not as a foundational paradigm [42] ? My own guess is therefore different from Hodges' : Turing *never gave up* his interest in his negative 1936 result concerning the uncomputable, as his 1949 article amply testifies [43], and *that is a reason why* he became interested in morphogenesis around 1950. Why should he link the uncomputable and morphogenesis ? Because the *emergence* of discrete natural forms was supposedly interpreted by Turing as an *effect* of the uncomputability of physical processes in nature [44] and this effect was scientifically unaccounted for at that time. The computable framework was therefore, to Turing's own mind, *not powerful enough* to tackle the general morphogenetic problem of how it is possible for a discrete form to emerge from a shapeless continuous substratum through symmetry breaking, which is the theme of his 1952 article. And this is *precisely* why Turing mentions the 'butterfly effect' in relation with the imitation game, since the discrete-state-machine model is *built* in such a way as *not to take* this effect into account. To repeat again Turing's own words [45] :

"The displacement of a single electron by a billionth of a centimetre at one moment might make the difference between a man being killed by an avalanche a year later, or escaping. It is an essential property of the mechanical systems which we have called 'discrete-state machines' that this phenomenon does not occur."

The impossibility of a mechanisable solution to the decision problem has two consequences : the tremendously powerful discrete-state-machine model *together with* its inner limitation, over which Turing never ceased to ponder [46]. That is why the formalist interpretation of the imitation game tells only half of the story. This will become even clearer if we go back to the imitation game itself.

2. Counter clockwise determinism : some arguments against the formalist viewpoint

I would like to show now that the formalist interpretation cannot make up the conditions expressly stated by Turing in the imitation game. To begin with, one must answer the question : are the experimental conditions devised in the game the same as those in a formal theory ? I have every reason to doubt it, as I will show now by presenting four anti-formalist arguments which all have to do with Turing's main point, i.e. the substitution of a Laplacian discrete-state machine for a man in game #2.

21. The return of the geometrical repressed

The first argument I will put forward is not mine : it was devised by G. Longo [47] and has to do with what was repressed throughout 19th- and 20th- century foundational research in the mathematical domain - the notion of a geometrical form as a purely *physical* phenomenon. The argument stands as follows : take game #2 played between a woman and a discrete state-machine and, after a while, let the interrogator ask each player to go back to its initial conditions

and start playing again. The difference between a human being and a computer will be immediately apparent because the computer, as a Laplacian determinist machine, will be able to go back exactly to its initial conditions and play exactly the same game, while this will *never* be the case for a human being or for most physical systems. In the case of a physical system - whether human or not - the idea of going back to exactly the same initial conditions is *meaningless* because they depend on a measure which is only determined within a certain interval. If a perturbation is small enough to be initially non-measurable, the system evolution may be subject to a 'butterfly effect', i.e., may become unpredictable. This is precisely what lies at the core of complex systems : the study of *geometrical structures* generated by non stable systems, just as in Turing's 1952 prophetic paper on morphogenesis. Therefore, the machine will be able to go in a loop and, starting from its initial conditions all over again, will repeat its evolution whereas a human being will not, disclosing its physical nature by this non-repetitive factor. Turing knows of course that this sensitivity to initial conditions is the reason why the brain is a "continuous system" and cannot be identified to a discrete-state machine [48] :

"The nervous system is certainly not a discrete-state machine. A small error in the information about the size of a nervous impulse impinging on a neuron, may make a large difference to the size of the outgoing impulse. It may be argued that, being so, one cannot expect to be able to mimic the behaviour of the nervous system with a discrete-state system."

So if intelligence is defined as a way of escaping the interrogator's decision, the computer will not be able to fool the interrogator, unless it succeeds in perfectly simulating a human being's behaviour. Turing precisely mentions this possibility when he imagines a dialogue between an interrogator and a machine in which the machine makes human 'mistakes' in order to fool the interrogator [49] :

"Question : Add 34957 to 70764
Answer : (Pause about 30 seconds and then gives an answer) 105621."

It is easy to check that 34 957 plus 70 764 does not make 105 621 but 105 721. Obviously, it looks like a careless mistake : the machine "forgot" to carry over the hundreds properly. This exemplifies the strategy of the machine : it must hide its superiority in arithmetic by introducing approximate results which look like careless mistakes, i.e., the kind of mistakes human beings are very likely to make. But is this enough to fool the interrogator definitely ? I do not think so, for at least one mathematical reason. Introducing some randomness in a determinist program is *not enough* when one wants to simulate dynamical systems with chaotic attractors. It has been rigorously proved that the organisation underlying these dynamical systems would be *destroyed* by a discrete-state simulation and therefore can *in no way* be simulated by a discrete-state program, even endowed with a randomizing feature- their very specific mixture of deterministic organisation and unpredictability lies therefore *beyond* any discrete-state simulation. What if a strategy appears in the game that would be best described as a chaotic behaviour of this kind ? In this case, what would be at least certain from the interrogator's point of view is that it cannot originate from a computer. This would therefore reinforce the interrogator's conviction that two human beings are actually playing. Taking the decision to quit from game #1 and begin playing game #2 would become even harder, if not impossible. This impossibility will become clear in my second argument which has a more logical aspect.

22. Logical undecidability in the imitation game

The actual setting of a dialogue in the imitation game depends upon an *imaginary* point of view which requires from the reader to be able to *recognize* the physical difference between a human being and a computer and at the same time *not recognize* this very difference. Let me explain this very crucial point. Although it is possible to conceive of the computer without any *petitio principii* under a double viewpoint, the first one being material (the computer is a collection of plastic parts, electrical lines and silicon) and the second one being non-material (the computer is a discrete-state machine), this cannot be the case for human beings because, in order for the imitation game to be convincing, *every human being reading the article has to take the point of view of the fooled interrogator while being at the same time able to make the physical difference between a human being and a computer*. The imitation game tries therefore to convince the reader that the physical difference between a human being and a computer is at the same time relevant and irrelevant, according to the point of view the reader chooses to adopt, either inside the game (in which, because the reader must identify with the fooled interrogator, the physical difference between the human being and the computer is abolished) or outside the game (in which the physical difference between a human being and a computer is given). It is the very possibility of this interplay between the inside and the outside of the game - i.e., this *undecidable* concerning the physical difference between a human being and a computer - which is never argued as such by Turing and which presupposes that the formalist distinction between the hardware and software *is already acquired* in the case of human beings, just as it is the case for computers. But this *was precisely what was to be experimentally established* and not only presupposed. That is why this point of view, at the same time inside and outside the game, is only *imaginary* and can never become formal. The fact that the imitation game can be played for real - as it is nowadays - does not change anything to this situation : the imaginary point of view is still necessary for the game to reach the goal it was meant for. The conclusion is therefore the following : in order to leave aside what pertains to the physical among the players and adopt a purely formal level of description, one has to adopt an *imaginary point of view*. As a matter of fact, I guess A. Hodges was on the verge of discovering this logical flaw when he wrote [50] :

"A deeper problem is that Turing's gender-guessing analogy detracts from his own argument. In the gender game, successful fooling of the interrogator proves nothing about the reality behind the screen. In contrast, Turing wants to argue that the successful imitation of intelligence is intelligence".

If game #1 does not prove anything about the definition of intelligence (or only that gender-difference cannot be detected through the imitation game) but if on the contrary game #2 proves something, it must be because of the presence of the computer. What difference does this presence make between the two games ? As I said, the computer can be seen from two points of view, i.e. as a physical piece of machinery and at the same time as a formal discrete-state machine. Game #2 is devised to generalise this dual point of view to human beings. But this implies adopting the *imaginary* point of view when the physical level of description is at the same time relevant outside the game and irrelevant inside it because we, as readers of the article, *are just human beings*, Matter and Mind intermingled. And this is why game #2 is concerned with the definition of intelligence and game #1 is not : game #2 *presupposes* that the notion of intelligence is a *formal* concept, entirely severed from any physical substratum. This will become even clearer if we look at the temporal limits of a game.

23. Physical undecidability in an imitation game

Turing defines two temporal limits, the first one inside the game and the second one outside it. The first temporal limit defines the duration of a game : 5 minutes ; the second one defines the duration during which a computer will become less and less detectable : 50 years. One point has to be made clear right away concerning the first temporal limit, i.e. the duration of a game, because it could induce a misinterpretation : one cannot see how a *finite* duration (5 minutes, 10 minutes or whatever duration is chosen before starting the game) could lead to the conclusion that the interrogator will *never* be able to do the right identification. Between a finite length of time and an infinite length of time, nobody would think that a finite length of time is enough to decide that the interrogator would *never* ask a revealing question. But I do not think this is Turing's argument - a five-minute game is in fact a *compromise* between the chances of success of the interrogator and the chances of success of the players. The duration of a game should be fixed in such a way as to let the game start but also as to let the players not too much time to play, otherwise the chances that they - and especially the computer - will be discovered get higher [51]. As for the second temporal limit, Turing's argument runs like this : if, in 1950, the chances of success on the interrogator's part are 100% and that they go down to 70% in 2000 and if the progression of failures tends towards 50% (i.e. if the interrogator's decisions are taken at random), then the time necessary to defeat the interrogator is about 50 years. The second aspect is this : at what time does it become suitable to decide the replacement of the man by a computer in game #1 ? One has to imagine that an external individual (someone watching the game or a potential reader of the article) who, having found out that the interrogator will not discover gender difference in game #1, decides to replace the man for a computer, thus changing game #1 into game #2. But on what basis can this external individual decide that game #1 has been played long enough, i.e. that gender difference is forever hidden from the interrogator's view ? The conclusion seems to me unescapable : if the external individual comes to this decision, it is because he has a preconceived opinion which cannot derive from the way the interrogator has asked some questions up to this moment. Where does it come from then ? It comes from the preconceived idea that there is a clear cut discrepancy between the physical and the intellectual. But this idea, at this stage of the argument, is not supposed to be shared by the reader who has to stick to the only criterium at his disposal - the verbal expressions of players are supposed to be sufficient to transform game #1 into game #2. Once again, the imitation game setting fails to reach its goal. This is quite clear when the player's strategies are under focus.

24. Gender-difference in the imitation game

The game has the ambition of showing that the level of intelligence dealing with imitation is absolutely disconnected from the level of biology dealing with gender difference. Contrary to what the explicit goal of the game is supposed to establish, the players' strategies do not have the same value and in fact, an implicit scale going from the less to the more abstract is being hinted at when a real game is going on. Turing's description of the woman's strategy is, from this point of view, quite revealing : "The best strategy for her is probably to give truthful answers. She can add such things as 'I am the woman, don't listen to him !' to her answers, but it will avail nothing as the man can make similar remarks."

It is plausible that one of the players should speak the truth in the game so that the other one can imitate his or her responses. *But there is no reason why telling the truth should be the woman's part exclusively* since it would be easy to imagine that the same strategy be played by the man. On the contrary, in Turing's description, the man's strategy is more of a real strategy than the woman's for it is a genuine imitation - that of the woman's responses. What about the

machine's responses ? It must be remembered that the machine replaces the man in game #2 - it is the man who is thus identified with a machine. As we have seen before, the main handicap of the machine is that it should manage to hide its "deadly accuracy" in arithmetic by introducing faulty results which look like careless mistakes, i.e., the kind of mistakes human beings are very likely to make. In this case, the machine does not imitate a particular gender of human beings but humanity. From this point of view, the machine is the only one to overcome the particular case of gender while imitating the only too human way of thinking. The grading in the responses becomes quite clear : starting from the woman's absence of strategy (she can imitate nobody but herself), it goes up to the man's strategy (he imitates the woman's responses) and then to the ultimate strategy, that of the machine replacing the man (it can imitate the human mistakes of whatever gender). But this scale rests upon a single, *unexplained*, fact - the absence of strategy on the woman's part. Thus one point must be acknowledged : if it is the woman who, for no apparent reason, is left aside in order to get rid of gender difference, then gender difference is not cancelled by the rationale of the game and plays a secret role in suggesting that displaying a computer-like strategy is a man-only job. *Therefore, gender difference is only denied but does not disappear in the game.* On what grounds could the woman's absence of strategy be justified ? My guess is that the imaginary point of view, necessary to reach explicit goal of the game, is not only logical and physical but also *sexual* since the imitation game is linked to a *fantasy*, that of abolishing once and for all gender difference in order to introduce intelligence as a formal concept. We saw that one of the main features of a computer was that, as a determinist discrete state machine, it was not sensitive to initial conditions. But this is precisely not the case with human beings since what makes two human beings diverge and become either a man or a woman is the introduction of a sexual difference as an initial condition. *The desire to abolish gender difference is therefore linked to the desire of building a physical object which would not be subject to the initial divergence of physical systems.* That is why the imitation game connects sexuality and the building of a discrete state machine. I showed elsewhere [52] that this fantasy is at the core of the invention of this most revolutionary mathematical and physical tool, which is the computer. This invention cannot be understood without referring to Turing's own fantasies whose recollections were made possible by the immensely valuable biography written by A. Hodges [53], although himself rather reluctant to such a psychoanalytical approach [54]. One last point should be stressed in following this trend of thought. I take it as an extraordinary feature how prophetic the imitation game is when one is aware of Turing's tragic end. Turing paid a very high price indeed to become aware that the initial sexual conditions of a human being cannot be changed at will without changing the human being who carry them throughout its own history - his suicide in 1954, following his chemical castration after being sentenced for homosexuality, should remind us that the idea of actually making someone change from that point of view can be psychically lethal. What is puzzling with Turing's sentence of 1952 is that it actually puts into practice this change in initial conditions which the imitation game in 1950 tried to establish as a possibility by imagining the replacement of a man for a machine. It looks as if the imitation game foreshadowed Turing's own fate, repressing in the undecidable conditions of the game an imaginary point of view which, one day, will *definitely* recapture him.

Conclusion

The imitation game, if it is studied carefully, is therefore much more intriguing and interesting than what the formalist interpretation pretends. It is not an opening to the research program of strong Artificial Intelligence, even if it can be misleadingly read this way. It is more of a global

meditation on the initial conditions of creativity, which doesn't solve the enigma but deeply contributes to its understanding.

Jean Lassègue

Laboratoire
Ecole Normale Maurice
1 rue Maurice
F-92120
France

Lattice-CNRS
Supérieure
Arnoux
Montrouge

jean.lassegue@ens.fr

Bibliography and web sites

Boden, M. (1990). "Introduction", *The Philosophy of Artificial Intelligence*, M. Boden, ed., Oxford : Oxford University Press.

Brouwer, L.E.J. (1913). "Intuitionism and Formalism", *Bull. Amer. Math. Soc.*, 20 : 81-96.

Cantor, G. (1932). *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts. Mit erläuternden Anmerkungen sowie mit Ergänzungen aus dem Briefwechsel Cantor-Dedekind*, E. Zermelo ed., Berlin : Springer.

Church, A. (1937). "Review of Turing 1936". *The Journal of Symbolic Logic*, (2) : 42-43.

Dupuy, J.-P. (2000). *The Mechanization of the Mind : on the Origins of Cognitive Science*, Princeton, N.J. : Princeton University Press.

Fefferman, S. (1987). "Weyl vindicated : "Das Kontinuum" 70 Years Later". *Temi e Prospettiva della logica e della filosofia della scienza contemporanea*. Bologna : CLUEB : 59-93.

French, R. M. (1990), "Subcognition and the Limits of the Turing Test", *Mind*, 99 : 53-65.

Gandy, R. (1988). "The Confluence of Ideas in 1936". *The Universal Turing Machine ; a Half-Century Survey*. R. Herken ed., Oxford, Oxford University Press : 55-111.

Girard, J.-Y. (2001). "Locus Solum". *Mathematical Structures in Computer Science*, vol.11, n° 3.

Gödel K. (1931). Über formal entscheidbare Sätze der Principia Mathematica und verwandter Systeme, Monatshefte für Math. Physik, 38 : 173-198 ; reprint in Kurt Gödel Collected Works, I, 1986, Cambridge University Press, Cambridge : 144-195.

Guillaume, M. (1994). "La Logique Mathématique en sa jeunesse ; Essai sur l'Histoire de la Logique dans la première moitié du vingtième siècle". *Development of Mathematics 1900-1950*. J.-P. Pier ed., Basel : Birkhäuser Verlag.

Hacking, I. (1975). *The Emergence of Probability*, Cambridge : Cambridge University Press.

Hallett, M. (1994). "Hilbert's Axiomatic Method and the Laws of Thought", *Mathematics and Mind*, A. George ed., Oxford : Oxford University Press : 158-200.

Haugeland, J. (1985). *Artificial Intelligence, the Very Idea*, Cambridge, Massachusetts : MIT Press.

Hilbert, D.

► (1899). *Grundlagen der Geometrie*. Berlin : Teubner.

► (1917). "Axiomatisches Denken". *Mathematische Annalen* 78 (1918) : 405-415 ; online at : <http://134.76.163.65/servlet/digbib?template=view.html&id=29014&startpage=409&endpage=419&image-path=http://134.76.176.141/cgi-bin/letgifsfly.cgi&image-subpath=/1408&image-subpath=1408&pagenumber=409&imageset-id=1408>

► (1922). "Die logischen Grundlagen der Mathematik". *Mathematische Annalen* 88 (1923) : 151-165
online at : <http://134.76.163.65/servlet/digbib?template=view.html&id=37001&startpage=155&endpage=169&image-path=http://134.76.176.141/cgi-bin/letgifsfly.cgi&image-subpath=/1571&image-subpath=1571&pagenumber=155&imageset-id=1571>

► (1925). "Über das Unendliche". *Mathematische Annalen* 95 (1926) : 161-190 online at : <http://134.76.163.65/servlet/digbib?template=view.html&id=26816&startpage=165&endpage=194&image-path=http://134.76.176.141/cgi-bin/letgifsfly.cgi&image-subpath=/1355&image-subpath=1355&pagenumber=165&imageset-id=1355>

Hodges, A.

► (1983). *Alan Turing ; The Enigma*. London : Unwin Paperbacks.

► (1988). "Alan Turing and The Turing Machine". *The Universal Turing Machine*, R. Herken ed., Oxford : Oxford University Press : 3-15.

► (1997). *Alan Turing ; a Natural Philosopher*. London : Phœnix.

Hofstadter, D. & Dennett, D. (1981). *The Mind's I*, New York : Basic Books.

Kant, E. (1790) *Kritik der Urteilkraft*. online at : <http://www.gutenberg2000.de/kant/kuk/kuk.htm>

Lachterman, D. R. (1989). *The Ethics of Geometry, a genealogy of modernity*, New-York and London : Routledge.

Lassègue, J.

- (1993). "Le test de Turing et l'énigme de la différence des sexes". *Les contenants de pensée*. D. Anzieu et al. Paris, Dunod : 145-195.

► (1996). "What Kind of Turing Test did Turing Have in Mind ?". *Tekhnema ; Journal of*

Philosophy and Technology, 3 : 37-58. (Online at : <http://tekhnama.free.fr/contents3.html>)

► (1998a). "Turing, l'ordinateur et la morphogenèse", *La Recherche*, 305, jan 1998, : 76-77.

► (1998b). *Turing*. Paris, Les Belles Lettres.

► (2003). "De la logique du mental à la morphogenèse de l'idée", dossier "Les constructivismes", M.-J. Durand ed., *Intellectica*, Paris ; (Online at : <http://www.ltm.ens.fr/chercheurs/lassegue/travaux.html>)

Leiber J. (1991). *An Invitation to Cognitive Science*, Oxford : Basil Blackwell, 1991.

Longo, G.
(Downloadable at <http://www.di.ens.fr/users/longo/download.html>)

► (1999). "The Difference between Clocks and Turing Machines", *Functional Models of Cognition*, Boston : Kluwer.

(2001). "Laplace" in "Locus Solum" (see Girard J.-Y.)

(2002). 'The Reasonable Effectiveness of Mathematics and its Cognitive Roots', *New Interactions of Mathematics with Natural Sciences* (L. Boi ed.), Springer.

(2003 to appear). "Laplace, Turing et la géométrie impossible du jeu de l'imitation", Paris : *Intellectica*.

Mc Culloch W.S. & Pitts W. H., (1943). "A logical calculus of the Ideas Immanent in Nervous Activity", *Bulletin of Mathematical Biophysics*, 5 ; reprint in Mc Culloch, W. S. (1965). *Embodiments of Mind*, Cambridge, Mass : MIT Press.

Michie, D. (1974). *On Machine Intelligence*. New York, John Wiley and Sons.

Peano, G. (1899). *Arithmetices principia, nova methodo exposita*, Augustae Taurinorum, Bocca.

Penrose, R. (1989). *The Emperor's New Mind ; Concerning Computers, Minds and The Laws of Physics*, Oxford : Oxford University Press.

Poincaré, H. (1913). *Dernières pensées*, Flammarion, Paris.

Pylyshyn, Z. (1984). *Computation and Cognition ; Toward a Foundation for Cognitive Science*, Cambridge, Mass : MIT Press.

Riemann, B. (1854). *Über die Hypothesen, welche der Geometrie zu Grunde liegen*, *Abhandlungen der königlichen Gesellschaft der Wissenschaften in Göttingen*, vol 13 : 133-150 ; online at : <http://www.emis.de/classics/Riemann/Geom.pdf>

Stewart, J. (2000). "Teleology in Biology", *Tekhnema ; Journal of Philosophy and Technology*, 6 : 14-33.

Turing, A. M.
- (1936). "On Computable Numbers with an Application to the Entscheidungsproblem". *Proceedings of the London Mathematical Society* 42 : 230-265 ; reprint in *Collected Works of*

A. M. Turing, vol. 4, Mathematical Logic, R. O. Gandy and C. E. M. Yates eds., Amsterdam, North-Holland ; online at : <http://www.abelard.org/turpap2/tp2-ie.asp>

► (1939). "Systems of Logic based on Ordinals". *Proceedings of the London Mathematical Society* 45 (ser 2) : 161-228 ; reprint in *Collected Works of A. M. Turing, vol. 4, Mathematical Logic*, R. O. Gandy and C. E. M. Yates eds., Amsterdam, North-Holland.

► (1949). "The Word problem in Semi-Groups with Cancellation", *Ann. of Math.* 52 (2), pp 491-505 (1950) ; reprint in *Collected Works of A. M. Turing, vol. 2, Pure Mathematics*, J. L. Britton ed., Amsterdam, North-Holland.

► (1950). "Computing Machinery and Intelligence", *Mind*, LIX, 236, Oct. 1950, 433-460 ; reprint in *Collected Works of A. M. Turing, vol. 1, Mechanical Intelligence*, Darrel Ince ed., Amsterdam, North-Holland : 133-160 ; on line at : <http://www.abelard.org/turpap/turpap.htm>

► (1952). "The Chemical Basis of Morphogenesis". *Phil. Trans. Roy. Soc. B* 237 : 37-72 ; reprint in *Collected Works of A. M. Turing, vol. 3, "Morphogenesis"*, P. T. Saunders ed., Amsterdam, North-Holland : 1-36.

Wagner, P. (1994). *Machine et pensée : l'importance philosophique de l'informatique et de l'intelligence artificielle*, Doctoral Thesis, Université Paris 1.

Weyl H, (1918). *Das Kontinuum, Kritische Untersuchungen über die Grundlagen der Analysis*, Veit Verlag, Leipzig.

Notes

[1] In (Turing 1950), Turing uses only three times the word 'test', two pages apart (446-447), while referring to the imitation game and answering the objection called "Argument from Consciousness".

[2] (Dupuy 2000).

[3] (Mc Culloch & Pitts 1943).

[4] To name just a few : (Michie 1974), (Hofstadter & Dennett 1981), (Pylyshyn 1984), (Haugeland 1985), (Penrose 1989), (Boden 1990), (French 1990), (Leiber 1991) ; The 'Turing Test' is even now an entry in the Collins dictionary : "Turing test - A proposed test of a computer's ability to think, requiring that the covert substitution of the computer for one of the participants in a teletype dialogue should be undetectable by the remaining human participant."

[5] Laplace is a useful reference point but Newton is of course the great figure.

[6] (Longo 2002).

[7] (Girard 2001).

[8] Other works are of course of great value. See for example (Hacking 1975) and also (Stewart 2000).

[9] (Longo 2001).

[10] It was only after the set-theoretical turn that it became clear that it was necessary to determine whether a given function was computable or not.

[11] (Longo 1999).

[12] G. Longo draws an analogy between Laplacian and Hilbertian determinism (respectively) in physics and in axiomatics on the one hand and Poincaréian and Gödelian-Turingian inner limitations of determinism in physics and in axiomatics on the other hand.

[13] (Feferman 1987) (Gandy 1988), (Guillaume 1994).

[14] (Frege 1884).

[15] (Peano 1889).

[16] (Hilbert 1899).

[17] (Hilbert 1917).

[18] a.k. their completeness (can every formula be either proved or refuted from the axioms ?), their consistency (can no contradictory formula be generated from the axioms ?) and their decidability (is there an effective way to always decide whether a formula is deducible from the axioms or not ?).

[19] (Brouwer 1913).

[20] See the Cantor / Dedekind correspondence in (Cantor 1932).

[21] (Poincaré 1913).

[22] (Weyl 1918).

[23] (Riemann 1854).

[24] Compared to Frege who was, as it were, a fundamentalist concerning the matter of reductionism, Hilbert himself was more of an agnostic : for him, the more interpretations a formal knowledge was able to receive, the better, because formal knowledge was 'reference-free'. On the contrary, Frege was claiming the absolute unicity of reference. Cf. (Hallett 1994).

[25] The continuum problem wrongly advertised as solved by Hilbert in (Hilbert 1925) in a proof he never completed, was the acid test for the reduction of higher-order infinity to finitary combinatorial propositions.

[26] (Lachterman 1989).

[27] (Hilbert 1923), p. 160 : "Unser Denken ist finit ; indem wir denken, geschieht ein finiter Prozeß."

[28] See for example (Hilbert 1925).

[29] (Gödel 1931).

[30] (Turing 1936), § 3.

[31] The "computer" in this quotation is of course a human being in the course of a computation ; cf. (Turing 1936), § 9.

[32] (Church 1937).

[33] This time construed as a machine and not as a human being performing a computation.

[34] (Turing 1950), § 5.

[35] This paragraph amply uses the argument I used in (Lassègue 1996).

[36] (Turing 1950), § 1.

[37] (Turing 1936), § 8.

[38] This interpretation was described by P. Wagner in (Wagner 1994).

[39] (Hodges, 1988), p. 9.

[40] (Turing 1938).

[41] (Hodges 1997), p. 28-29.

[42] (Lassègue 1998a).

[43] (Turing 1949).

[44] Very much like what Kant did in *Kritik der Urteilskraft*, § 78 : "[...] so wissen wir auch nicht, wie weit die für uns mögliche mechanische Erklärungsart gehe, sondern nur so viel gewiß : daß, so weit wir nur immer darin kommen mögen, sie doch allemal für Dinge, die wir einmal als Naturzwecke anerkennen, unzureichend sein."

[45] (Turing 1950) § 5.

[46] A. Hodges is of course fully aware of this, since, after quoting Turing's very sentence, he comes to the conclusion : "Thus, we cannot feel that Turing had arrived at a complete theory of what he meant by modelling the mental functions of the brain by a logical machine structure. But we should give proper credit for raising such questions at all." (Hodges 1988), p. 11.

[47] (Longo 2003). The main point of the article is to determine the scope of Laplacian determinism as well as the possibility of rationally expressing its inner limitation.

[48] (Turing 1950) § 6, (7).

[49] (Turing 1950) § 2. The mistake was noticed by D. Hofstadter. But he only mentions it and does not seem to take any advantage of this crucial fact ; see (Hofstadter & Dennett 1987), p. 667-668. (Turing 1950) mentions also this strategy more explicitly in § 6 (5) : "It is claimed that the interrogator could distinguish the machine from the man simply by setting them a number of problems in arithmetic. The machine would be unmasked because of its deadly accuracy. The reply to this is simple. The machine (programmed for playing the game) would not attempt to give the right answers to the arithmetic problems. It would deliberately introduce mistakes in a manner calculated to confuse the interrogator."

[50] (Hodges 1997) p. 38.

[51] This is typically a problem of sequential analysis Turing became acquainted with during World War II.

[52] See also (Lassègue 1993), (Lassègue 1996) and (Lassègue 98) p. 188-189.

[53] (Hodges 1997).

[54] (Hodges 1997) p. 38 : "But if Turing's gender-game is misunderstood, he certainly courted such confusion. He panted the pages of this journey into cyberspace with the awkward eroticism and encyclopaedic curiosity of his personality. Modern cultural critics have jumped with delight to psychoanalyse its surprises. The intellectual text is the austere statement of the capacity of the discrete state machine for disembodied intelligence, the subtext is full of provocative references to his own person, as if putting his own flesh-and-blood intelligence on trial".