MICHAEL POLANYI: CAN THE MIND BE REPRESENTED BY A MACHINE?

Documents of the Discussion in 1949

PAUL RICHARD BLUM

Loyola University Maryland
prblum@loyola.edu

ABSTRACT

In 1949, the Department of Philosophy at the University of Manchester organized a symposium "Mind and Machine" with Michael Polanyi, the mathematicians Alan Turing and Max Newman, the neurologists Geoffrey Jefferson and J. Z. Young, and others as participants. This event is known among Turing scholars, because it laid the seed for Turing's famous paper on "Computing Machinery and Intelligence", but it is scarcely documented. Here, the transcript of this event, together with Polanyi's original statement and his notes taken at a lecture by Jefferson, are edited and commented for the first time. The originals are in the Regenstein Library of the University of Chicago. The introduction highlights elements of the debate that included neurophysiology, mathematics, the mind-body-machine problem, and consciousness and shows that Turing's approach, as documented here, does not lend itself to reductionism.

Keywords: artificial intelligence; computer; epistemology; formalization; mathematics; mind and machine; neurology; neurophysiology; philosophy of mind; Turing test

On the 27th of October, 1949, the Department of Philosophy at the University of Manchester organized a symposium "Mind and Machine", as Michael Polanyi noted in his *Personal Knowledge* (1974:261). This event is known, especially among scholars of Alan Turing,[1] but it is scarcely documented. Wolfe Mays (2000) reported about the debate, which he personally had attended, and paraphrased a mimeographed document that is preserved at the Manchester University archive. He forwarded a copy to Andrew Hodges and B. Jack Copeland, who then published it on their respective websites.[2] This document, minutes taken by an anonymous member of the Philosophy Department, will be published here with some footnotes intended to identify the persons present and the contexts in which they spoke. The basis of

[1] Hodges 1983:487; cf. Turing 2004:487 n. 1. In connection with Polanyi see Scott and Moleski 2005: 215, and Blum 2010a.

[2] See http://www.turing.org.uk/sources/wmays1.html with commentary by A. Hodges, and http://www.alanturing.net/turing_archive/archive/m/m15/M15-001.html, posted by B. Jack Copeland who also published a transcript in *The Rutherford Journal*, vol. 1, December 2005: http://rutherfordjournal.org/article010111.html.

© POLANYIANA 2010/1-2. (19): 35–60.

this publication is the copy preserved in the Regenstein Library of the University of Chicago, Special Collections, Polanyi Collection (abbreviated RPC, box 22, folder 19).[3] The same collection holds the mimeographed statement that Polanyi prepared for this symposium: „Can the mind be represented by a machine?".[4] This text has not been studied by Polanyi scholars and is also published here. Since one of the participants was the neurologist Geoffrey Jefferson, a brief piece inscribed „Jefferson's last lecture" that is among the Polanyi papers (RPC box 24, folder 1) and which addresses the question of mind and machines will also be published here.

First a summary of the debate as it appears from the minutes; an interpretation of Polanyi's statement will follow.

From the minutes of the discussion, it is obvious that the question concerning the human mind and computing machines was addressed—and should be addressed—from a variety of scientific angles. Indeed, the minutes could be read as testimony to the fact that computers as they are known and artificial intelligence emerged from a genuinely interdisciplinary debate and effort. And it would be easy to show that the antagonists at Manchester are representative of a cooperation, rather than opposition, to that end. However, here it seems to be conducive to present the argumentative threads separate by discipline, starting with philosophy. Dorothy Emmet, a scholar of Alfred North Whitehead, seems to have kept a low profile and yet she raised the typical philosophical question, whether consciousness is not what distinguishes human thought from machine operations. It is interesting that Alan Turing, in devising his later so called Turing Test, related this objection not to a philosopher but rather to the neurologist Jefferson (Turing 1950:446). We have no evidence from the notes whether Jefferson had agreed with Emmet, but Turing's answer reveals that his approach to 'machine thought' focused on „sets of rules"; he granted „conscious working" a status separate from routine operations that can be performed/emulated by a machine. Thus, looking for the philosopher's input first, we see immediately everyone concerned with the specter of materialism or of the „fourth continuity", i.e., the upcoming new shock (after Copernicus, Darwin, and Freud had evinced the continuity between humanity and the cosmos, animals, and mental illness) which consisted in creating a seamless transition between mind and machine (Mazlish 1967).

[3] I am grateful to the Special Collections Research Center, University of Chicago Library, for the permission to publish the documents of this paper. I am also indebted to the Center for the Humanities of Loyola University Maryland for a summer research grant and to the Jesuit community of South Chicago for their hospitality in June 2009. Jennifer Sanchez and Emmett Holman made valuable suggestions. The same RPC box 22, folder 7, also has a one page mimeographed statement by Alan Turing, which Jack Copeland is planning to publish in *The Rutherford Journal*. The text starts with the words: „We are interested in machines which test the description numbers (D.N.s) of other machines to see wh[e]ther they are 'satisfactory', i. e. whether the machines concerned print an infinity of 0s and Is." The text has no header, but it bears in Polanyi's hand the name „Turing".

[4] RPC box 32, folder 6; mentioned in Scott and Moleski 2005:215.

It seems Emmet was the only professional philosopher present, if we except Polanyi who in 1948 had moved from chemistry to „social studies".[5] Mathematicians were present, namely Maxwell Herman Alexander „Max" Newman, Alan Turing, Maurice Bartlett, and Bernhard Neumann; furthermore the neurologists J. Z. Young and Jefferson, and then a person named Hewell (so far unidentified; note the question mark at the first appearance of his name) who also seems to be an expert in physiology of the brain.

Evidently, the discussion started with Newman responding to Polanyi's statement and his interpretation of Gödel's theorem. It appears that the mathematician conceded the assumption that the operations of machines „cannot do anything radically new" and that the assessment of the difference between mind and machine is a matter of experimental research, rather than a priori speculation. At this point, Turing intervenes by emphasizing that his idea of a „universal machine" entails some kind of self-referential operation, as one could translate the capability of „turning itself into any other machine". Consequently, the three questions summarized by Emmet after a break: machine-brain analogy, physiology, limitations of the machine, are all answered by the mathematician Turing with an enigmatic reference to „trial and error" in combination with memory. Turing seems to have no doubt that operating on 'past experience' and memory are not features that keep the machine and the brain apart.[6] The next time a mathematician joins in, namely Newman, the problem of the design of the calculator comes to the forefront. He implicitly suggests using the design of a calculating machine, as the paradigm for investigating the mind. In doing so, he underlines the elementary problem raised by physiologist, Young, of internal versus external approach to operations. For „in the case of the mechanical brain we start with something which has been made by us", as Young said, so that the implied solution could be a methodical approach to what takes place when a calculating machine is being programmed. This would be a 'meta-programming' approach, in which the program and the design are not the same thing.

When the discussion circles around the problem of memory storage, Newman makes a further methodical suggestion that, broadly speaking, reflects the hypothetical-deductive method of science: „start like the atomists with a 'billiard ball' hypothesis"; that is to say, to hypothesize that the mind is a machine („which is obviously wrong") in the hopes that experimental and theoretical research will falsify that, in due course. Not only the mathematician Neumann, with reference to consistency proofs, but also the neurologists agree. In spite of Polanyi's repeated objection that some achievements of the mind cannot be hypothesized with „crude models", the conversation goes on along the lines of models, analogies, and hypotheses and heads towards the notion of „incompatibles". If a system is observed from the

[5] His position was „Social Studies" within Political Economy: Charlton 1951:175 and 177.

[6] Cf. on self-reference and memory in computer-intelligence Copeland and Proudfoot 2005:124.

outside, hypothetical paradigms help pointing out crucial changes that may or may not reveal alternatives or „choices" in the observed process. At this point, Turing stresses that „random operation can be made to become regular after a certain prevailing tendency has shown itself". The regularity is to be assumed first in the observation, that is, in the accommodation of the paradigm to the observed process, and then assumed to be indicative of the regularity of the process itself. So far it is a case of pure empirical research. Turing illustrates that with the operation in a machine, which on the input of incompatible data registers the contradiction and returns to the origin of the contradiction. Just recently, John von Neumann had suggested that the rounding-error of a calculator, which is caused by the limited number of digits, can be compensated by having three machines doing the same calculation and stopping in case of conflicting results.[7] This is the moment when Turing explicitly sides with Polanyi in citing his insistence on the basic difference between „mechanically following rules" and consciously knowing rules. And yet there remained a dissent that needs to be clarified. Before that can be done, we need to take a look at the physiological/neurological view in the debate.

To summarize the mathematicians' contribution to the discussion: there is no doubt that machines are no minds, although Turing is seeking for the option to produce self referring machines, which opens the methodological discussion about design, method, and the meaning of rules.

Early in the debate Jefferson feels compelled to state that bodily functions may be interpreted in mechanical terms, „but not 'Mind'", to which Turing replies that even in engineering and operating a machine, there remains an element of playing and ignorance. If that was actually the response of Turing to Jefferson, as it appears in the minutes, then the mathematician is consciously rejecting a mechanistic approach to physiology. His later definition of purpose as „use of previous combinations plus trial and error" suggests, in this perspective, that biologists should not abolish hypothetical purposes in living organisms, but rather redefine purposefulness in terms of programming. In terms of programming, to „put a purpose into a machine" is not different from the operations of living organism. Bartlett seems to have no problem assuming that the brain can be functioning with statistical errors. So Young joins in: first, he interprets Turing's remarks as an invitation to neurologists to collaborate with mathematicians. Cybernetics would be the point in case. (Turing was involved in the birth of this new discipline: Wiener [1948] 1961:23.) Then Young mentions as a problem that the behavior of brain cells might be different from that of other cells in organisms. But more importantly, he points out that the „collaboration" could reveal a fundamental difference between physiology that investigates a self-sustaining object, whereas engineering a brain provides an

[7] Von Neumann 1967:35; the Hixon Symposium on „Cerebral Mechanisms in Behavior" took place in September 1948, the papers were published in 1951.

object, the rules of which have been established and implanted by the researcher. He re-phrases the same difference by asking: „The physiologist can stimulate points and see what happens—do the 'mechanicians' do the same?" The distinction between a mechanical and a physiological view on the mind could be that the brain as an organism is still good for surprises (it has to be studied empirically); whereas the mind as a machine would presuppose that its laws are known a priori (from its blueprint). Nothing unexpected should happen when poking a mind-machine. Histology and EEG are mentioned as recent techniques of physiological investigation and the specialists keep exchanging about its uncertainties until Young moves over to philosophical aspects.

In departing from empirical approaches, the physiologist seeks help from philosophy: first Young ponders the option that memory is not limited to a specific location in the brain, like memory cells, and then he endorses the term Gestalt of philosophical psychology and calls for logic to set up promising hypotheses. The promise lies in the pattern shift. Wholeness,[8] paradigms, and logically supported hypotheses—that's what is needed in order to guide empirical research into the brain as the organ of thought. Polanyi very much appreciated Gestalt psychology giving it an anthropological and cognitive meaning: „Gestalt psychology has assumed that perception of a physiognomy takes place through the spontaneous equilibration of its particulars impressed on the retina or on the brain. However, I am looking at Gestalt, on the contrary, as the outcome of an active shaping of experience performed in the pursuit of knowledge."[9] From this philosophical interest he can voice doubt whether „seeing stereoscopically" can play a role in the envisioned research program on the mind and, again, how would is possible to „derive from the model the conception of 'seeing in depth'"? The philosopher seems to be capitalizing on the inherent methodical limitations of mechanical physiology, which are both expressed in Jefferson's naïve dualism and in the antinomies of organisms that, as empirical objects, elude a singular method.

Eventually the physiologists joined the mathematicians in departing from the narrow confines of their disciplines. Cybernetics, in this perspective, opens an understanding of machine operation that matches the anatomical findings and lack thereof. Jefferson (1960:43) lamented later that feelings miss „special abodes in the brain" and have been reduced to „fictitious … entities". Since he was quite informed about the history of sciences, and in particular the debate about the location of the soul (Jefferson 1949, cf. 1960:94-209), it is also legitimate to mention that the study of

[8] Piaget (1969:xxvi) defined Gestalt to be one of the „major concepts of wholes, … a structuralism without genesis".

[9] Polanyi 1967:6 (originally 1961). Cf. Polanyi 1974:55–61, and Mullins 2010. Interestingly, Douglass Hofstadter (1999:574) labels Polanyi „a holist par excellence" while presenting the Church-Turing Thesis on computers and the brain. For a physiological account of Gestalt see Wiener [1948g] 1961:139 and 141.

the physiology of the brain, in connection with attempts at anatomy and localizing the human mind and its functions, had had a steady career in 18th through 19th century medicine and philosophy (Hagner 2008). Mechanical optimism battled with philosophical skepticism. There was always at stake the question: is there any relation between functions of the human mind (sense perception, memory, morality, etc.) and the empirical data? Young's probing octopus brains and Jefferson's neurosurgery were just continuations of that century old debate. The aim remained the same, finding the interface between psychic and physical states. The news at this discussion was that decentralized memory storage and generally non-localized and non-mechanical forms of operation became thinkable. In that sense it marked the threshold to the computer age and cyber world.

As soon as unpredictability, randomness (Turing), non-quantifiability, and holistic approaches are admitted, Turing is able to redirect the perspective away from mind-as-a-machine towards the functioning of machines. Therefore, when he answers Jefferson's objection who doubted that human beings would be able to be perturbed by conflicting results of a thought process, Turing quips—to the amusement of the people present—that this is what mathematicians do. So we have the paradoxical situation that the physiologist gives less credit to the 'intelligence' of the mind than Turing gives to a well working machine. When someone in the audience asks: „are mathematicians human beings?" it becomes obvious that the Turing project is to analyze human thought by way of programming a computer. Mathematics is what humans do and what machines can do.

Therefore it now also becomes clear what is at stake between Turing and Polanyi. They agree that a mathematical interpretation of thought is not all there is. Yet, Turing tries to find in thinking as much mathematical procedure as possible, while Polanyi aims at capturing with philosophical precision that what remains. Jefferson's foggy „not the 'Mind'" and Young's apparently crestfallen swerving into philosophical methodology are to be remedied by Polanyi's insistence on formalization and specification. This is what he had to tell his colleagues in his own statement.

In his statement on the question whether the mind can be represented by a machine Polanyi pronounced five theses. First he interpreted the development of mathematics from Hilbert to Gödel as establishing a realm that cannot be formalized and hence is prior to computation that a machine can do. Second, he identifies this non-computing operation with reflection as the specific power of the mind. Third, the outcome of Gödel's and Tarski's discoveries do not disturb the understanding of human mind, they rather afford a philosophical tool to distinguish the primordial capability of reflection on rules which itself is not bound to those rules. His fourth point is the capability of belief that precedes empirical knowledge and is its foundation. Lastly Polanyi reasserts the denial of mechanical determinism. In Roger Penrose's classification (1994:12–16) Polanyi would probably be a „C- believer", according to whom „the problem of conscious awareness is indeed a scientific one, even if the appropriate science may not yet be at hand" (1994:16), that is, Polanyi seems to believe that some mental activities can be emulated by computers, but not all of them.

As is well known, Gödel had claimed that the system S contains propositions that cannot be proven and it contains undecidable problems (paraphrased from Gödel 1930:141–143). He had also added that his theorems „can be extended also to other formal systems" (Gödel 1930:143); indeed, „[a]ny epistemological antinomy could be used for a similar proof of the existence of undecidable propositions" (Gödel 1931:149 n 14). Now it is revealing how Polanyi phrased Gödel's „discoveries": To him this was originally about „arithmetic and advanced geometry". An unknown hand corrected his wording by saying that Gödel dealt with „number theory". This is factually correct, but it shows Polanyi's drive to extend the meaning of the theorem beyond number theory. A few lines down, when Polanyi concluded that there must be a „procedure for the discovery which, by its very nature, is incapable of formalization", the same hand interjected that formalization is possible in meta-language. For Gödel 'formalization' was a term of art within mathematics, that is, to be „reduced to a few axioms and rules of inference", and his aim of 1930/1931 had been to show that it is not the case „that these axioms and rules of inference are sufficient to decide any mathematical question that can at all be formally expressed in these systems" (Gödel 1931:145). This exchange makes it clear that Polanyi saw in Gödel's discoveries a point of departure from the need of formalization. As a scientist he was, of course, well acquainted with axiomatic systems. But as a philosopher, he was intrigued by the option of an infinite regress in formalization, a regress that is spurred by reflection, as he says, to the effect that reflection must necessarily stand outside of the mathematical/scientific procedure. When his reader appealed to meta-language that could reenter the process of formalization, i.e. axiomatization, he was kicking at an open door, for Polanyi had already integrated meta-language on his escape route out of the world of formalization. Tarski and Gödel are both witnesses to an „indefinitely extending programme of innovation, which can be achieved only by informal methods and not by a machine" (end of section 2).

When Polanyi returns to the subject in his *Personal Knowledge* he states (1974: 258) that:

> …a formal system of symbols and operations can be said to function as a deductive system only by virtue of unformalized supplements, to which the operator of the system accedes: symbols must be identifiable and their meaning known, axioms must be understood to assert something, proofs must be acknowledged to demonstrate something, and this identifying, knowing, understanding, acknowledging, are unformalized operations on which the working of the formal system depends.

Thus he groups a wide range of mental acts together as those that precede formalization (symbols, axioms, rules), and this is what in the 1949 debate he called „semantic function". „We call them the *semantic functions* of the formal system.

These are performed by a person with the aid of the format system, when the person relies on its use." (Ibid.) He immediately adds that there is a „legitimate purpose of formalization", namely, an increasing reduction of „informal operations; but it is nonsensical to aim at the total elimination of our personal participation" (259).[10]

So he remains faithful to his understanding that mechanization, and specifically a mathematical interpretation of thought, is appropriate as long as one acknowledges the existence of what he later would call the personal coefficient. Surprisingly he dismisses the Turing Test (Turing 1950) alleging that Turing had turned the question about thinking machines into „the experimental question, whether a computing machine could be constructed to deceive us as to its own nature as successfully as a human being could deceive us in the same respect" (263 n 1). If Polanyi did not reject Turing's project on the whole—and he didn't—he must have been alarmed by the playful implications of the mathematician's tongue-in-cheek approach to cognition. It is obvious that Turing's mental experiment successfully fooled a large audience into believing that thinking was a trickster game.

Much of Polanyi's energy was invested in unmasking imposters and simplifiers. In the opening chapter of *Personal Knowledge*, he takes to task the myth of objectivity that makes believe there were no personal investment in discovering objective facts of nature (cf. Blum 2010b). In a paper of 1950 on „Scientific Beliefs" (that will become part of *Personal Knowledge*¸ chapter 12), he attacked standard positivism: „A genuine scientific theory must operate like a calculating machine, which, once the keys representing the dividend and the divisor have been depressed, determines the result automatically" (27). It is this broader cultural context that interested Polanyi when he joined the debate about mind and machine. The misunderstanding of the working of a machine is an expression of the mistaken anthropology. Positivists and mechanicists believe to „construct a machine which will produce universally valid results. But universal validity is a conception which does not apply outside the commitment situation." (35). Polanyi dedicated an offprint of that paper „to A. M. Turing with best thanks".[11] In a later class on „Unspecifiable Elements of Knowledge" (his famous book was out since 1958) he boldly uses machines as a paradigm. When the design of a machine had been invoked to solve the problem of the mind-machine-riddle, then it was even more fitting to choose as the „leading example a class of comprehensive entities of which we can specify both the particulars and their coherence."[12] The surprising result is that the philosophy of machines is not much developed, and therefore it has been overlooked that machines „embody

[10] Cf. Polanyi 1952:312: „I think it is logically fallacious to speak of a *complete* elimination of what has been called 'psychological' but might better be called 'unformalised' elements of deductive systems ..."

[11] http://www.turingarchive.org/viewer/?id=512&title=main

[12] „Syllabus of class on Unspecifiable Elements of Knowledge, Hilary Term 1961", RPC box 22, folder 14. Quotations from „Fourth meeting".

rules that are not laws of nature"; even more, the failures of machines („bursting of boilers") are part of their essence, namely, as „imperfect embodiment of its ideal". In this lecture Polanyi fought submitting to the machine as the ideal or paradigm and recuperated it for the anthropological inquiry of knowledge.

Therefore, it is important to notice Polanyi informing his readership of *Personal Knowledge* that Turing's contribution to the Symposium „Mind and Machine" was „foreshadowed" by his paper on „Systems of Logic Based on Ordinals". That paper „deserves to be read and understood far more than it has been." (Turing 1939:71, introduction.) It addressed Gödel's incompleteness theorems, but towards the end, the author leaves technical mathematical language behind and reflects upon mathematical reasoning; and it was most likely this § 11 that caught Polanyi's attention. „Mathematical reasoning may be regarded rather schematically as the exercise of a combination of tool faculties, which we may call *intuition* and *ingenuity*." (Turing 1939:214) In a footnote he clarifies that he is „leaving out of account that most important faculty which distinguishes topics of interest from others; in fact, we are regarding the function of the mathematician as simply to determine the truth or falsity of propositions." Again an example of Turing's sense of irony: the most important thing is left out. But that leaves us with understanding that the reach of his number theory goes exactly as far as truth and falsity of propositions go. This is no reductionism. In describing the function of intuition and ingenuity, he emphasizes the role of intuitive judgment and the need for „suitable arrangements of propositions" and takes it for granted that „these two faculties differ of course from occasion to occasion, and from mathematician to mathematician." Again, assuming that Turing is leaving technical mathematical language behind, his description can only be understood as the establishment of the competence that Polanyi would have understood to be 'personal'. It seems this is the passage Polanyi had in mind when he invoked Gödel for having proven that there is a non-formalized capability of the human mind. „In pre-Gödel times it was thought by some that it would probably be possible to carry this program to such a point that all the intuitive judgments of mathematics could be replaced by a finite number of these rules. The necessity for intuition would then be entirely eliminated." (Turing 1939:215) Reductionism in the sense of eliminating the personal component has been overcome by Gödel and, consequently, Turing. To eliminate the personal component is a methodical aim for the sake of mathematical theory. Therefore during states: „We are always able to obtain from the rules of a formal logic a method of enumerating the propositions proved by its means. We then imagine that all proofs take the form of a search through with this enumeration for the theorem for which a proof is desired. In this way ingenuity is replaced by patience." He calls that „heuristic". (Ibid.)

Turing's paper on „Systems of logic based on ordinals" may have been extraordinary within his own production, but Polanyi found it more important than his famous 1950 paper. It seems the mathematician was „steaming straight ahead with the analysis of the mind, by studying a question complementary to 'On Computable Numbers'," as Andrew Hodges put it.

> The Turing machine construction had showed how to make all formal proofs 'mechanical'; and in the present paper such mechanical operations were to be taken as trivial, instead putting under the microscope the non-mechanical steps which remained. (Hodges 1999:19–21)

Therefore if the analysis of human thought is at the focus of attention, the distinction upon which Polanyi and Turing agreed, namely, that between rules and knowing rules turns out to be constitutive for any theory of thinking and computing.

As a young man, Turing had mused about the „Nature of Spirit" and described the same relationship as follows:

> As regards the question of why we have bodies at all; why we do not or cannot live free as spirits and communicate as such, we probably could do so but there would be nothing whatever to do. The body provides a something for the spirit to look after and use. (Hodges 1983:64)

This is patently the traditional language of body-soul-dualism, and it will take some education to translate that into problems of logic and mathematics. But looking back the structural identity is clear. There is a relationship of independence and manifestation that cannot be 'reduced' or 'eliminated'. Obviously pure non-formalized thought would be as 'boring' as an absolutely free spirit. On the other hand, science is after the laws of matter. Polanyi expressed that in the context of his recapitulation of the Manchester debate by assuming that a mechanical approach implies to determine that a particular object is seen to be a machine, a perspective that in and of itself leaves already simple mechanistic views behind.[13]

> A machine is an interpretation of an observed mind and not of an observing mind." (Polanyi 1952:315) „For a machine is a machine only for someone who relies on it for some purpose, that he believes to be attainable by what he considers to be the proper functioning of the machine: it is an instrument of a person who relies on it. (Polanyi 1974: 262)

[13] In this context it is impossible to follow the ontological implications of this perspective; therefore see Margitay 2010, specifically 135–138 on the question whether there is a difference in structural determination between artifacts and nature.

1. Discussion on the Mind and the Computing Machine[14]

These notes of the Discussion on October 27th were taken by a member of the Philosophy Department, and you may be interested to have them as a record.

Rough draft of the Discussion on the Mind and the Computing Machine, held on Thursday, 27th October, 1949, in the Philosophy Seminar.

NEWMAN[15] TO POLANYI: The Gödel extra-system instances are produced according to a definite rule, and so can be produced by a machine. The mind/machine problem cannot be solved logically; it must rest on a belief that a machine cannot do anything radically new, to be worked on experimentally. The interesting thing to ask is whether a machine could produce the original Gödel paper, which seems to require an original set of syntheses.

TURING emphasises the importance of the universal machine, capable of turning itself into any other machine.

POLANYI emphasises the Semantic Function, as outside the formalisable system.

JEFFERSON[16] will admit that the respiratory system is mechanical, but not 'Mind'.

TURING: One may 'play about' with a machine and get the desired result, but not knowing the reason; an element of this kind enters into both engineering and operating it.

Resumption at 6.40 p.m.

EMMET[17]: Questions to be considered.
(1) Machine-brain analogy;
(2) Physiological aspects;
(3) Are there any limitations to the kind of operations which a machine can do?

[14] RPC; box 22, folder 19. The document has no header. Another copy is online available at http://www.turing.org.uk/sources/wmays1.html with commentary by Andrew Hodges, and http://www.alanturing.net/turing_archive/archive/m/m15/M15-001.html; transcript by B. Jack Copeland in *The Rutherford Journal*, vol. 1, December 2005: http://rutherfordjournal.org/article010111.html.

[15] Max Newman (1897-1984), professor of pure mathematics at Manchester University since 1945 (Charlton 1951:176; Knowles 1959).

[16] Geoffrey Jefferson (1886-1961), professor of neurological surgery at the University of Manchester since 1939 (Charlton 1951:181). This debate is not documented in the otherwise detailed biography Schurr 1997 (p. 271 covers at that time); Jefferson had just delivered his talk on „Descartes and the localization of the soul" and his lecture „The mind of mechanical man" (the Lister oration).

[17] Dorothy M. Emmet (1904-2000) was a philosophy professor at the University of Manchester (Knowles 1959), she organized this event.

Questions asked: Is it possible to give a purpose to a machine? Can you 'put a purpose into a machine?'

TURING: This kind of thing can be done by 'trial and error' methods: purpose is 'use of previous combinations plus trial and error'.[18]

BARTLETT: Even in the ideal calculating machine you have a small statistical error, the latter we find also in the brain.[19]

YOUNG: Speaking from the physiological point of view: he is looking at the point from a purely practical point of view. He feels that none of the collaboration which has so far taken place has enabled him to ask the right questions. Neuro-physiology is not progressing: Cybernetics[20] is based on 'Models' by which we work. Another point: do agglomerations of brain-cells act in the some way as individual ones do in the other parts of the body? There seem to be chains which are not functioning as conductors only, as has been hitherto thought. But if this is so—how should we proceed?

No one knows just what to measure, and it is in the ordering of the attack that some collaboration might result. The physiologist starts with a system not made by himself, but in the case of the mechanical brain we start with something which has been made by us.[21] Is the approach then identical? If not, can the right approach be suggested? The physiologist can stimulate points and see what happens[22]—do the 'mechanicians' do the same?

NEWMAN: Possible approach: it might be asked how the calculating machine was designed - approaching the thing from the outside as it were. Could methods used in answering this question be applied to the other?

[18] „Turing was obviously thinking here of feedback mechanisms, sometimes called goal-seeking devices, which by trial and error gradually approach the target or goal. However, If [sic] such a device can be said to have a purpose it is only because it has been programmed into it by a human being." Mays 2000:62.

[19] Maurice S. Bartlett (1910-2002) professor of mathematical statistics in Manchester since 1947 (Charlton 1951:176 and Knowles 1959). Cf. obituary Gani 2002.

[20] „Cybernetics" was a new term, dating back to 1947 and made public by Norbert Wiener ([1948] 1961:12) in his „entertaining book" (Jefferson 1949:1108) by that title. The development of the discipline involved, among others, Turing, whom Wiener met in Manchester in 1947, and von Neumann and McCulloch (Wiener [1948] 1961:15 and 23; Hodges 1983:413 f.). Young (1971:99) linked cybernetics with homeostasis, i.e. the capability of living organisms to maintain their unsteady state (i.e., survive): „By making machines that imitate and assist one of our most important functions we have once again learned to speak more precisely about that function in ourselves. This is a part of the science known as cybernetics [...] (Wiener 1949). [...] The essential feature of that organization is that it maintains itself intact in spite of changes in the surroundings. This homeostasis is made possible because the organism receives information aubout the changes in its environment and this information controls the course of its actions. We can therefore speak of the whole process by comparing the body with machines whose course of action is controlled by information received from some outside source."

[21] Cf. end of previous quotation: the exteriority of the information is necessary for both, mind and machine.

[22] Young 1971:620: Diagram of stimulations in brain cells.

……….?: replies that impulses could also be applied to the machine to see what happens in this case.

JEFFERSON: claimed that histological investigations which give visual data are easier.

YOUNG: The E.E.G, gives interesting elements—they do not depend on 'circuit' conceptions.[23]

JEFFERSON: E.E.G. gives a very general result (produced diagrams). The exposed brain gives us no better trace than an intact one still in the skull. The E.E.G. has therefore a certain use, but it is not a fine enough method.

YOUNG: disagrees, since we are dealing with cells in large groups.

JEFFERSON: points out that if a man is set a problem e.g. to multiply 13 by 17, then the trace stops.

YOUNG: suggests that it might be of use on scanning large areas of the brain, with the aim of finding then an assumption applicable to each individual cell.[24]

HEWELL: (?) (I.C.I. Research)—dealt with the analysis of E.E.G. traces; pointed out that it is possible to tell in which stage of a fit a patient finds himself by analysing the E.E.G. records. The Electrocardogram [sic] gives a typical waves-form from the heart, and if the latter is reduced to one small piece of fibre, the trace continues even when there is no beating motion. Hence we have something which is passing over the organism, and not something in it. (?).

JEFFERSON thinks that E.E.G. is not much use.

YOUNG disagrees.

JEFFERSON: E.E.G. trace-vibration depends on many factors (blood-sugar etc.) hence we are never sure when we are dealing with a 'normal' subject.

YOUNG thinks comparative method, either between people, or between animals, would be useful.

HEWELL: if certain areas (two of them) of the hypothalamus are stimulated, we have then a certain reaction ('petit mal') which leads us to the supposition that 'scanning' is going on in the brain.

[23] Norbert Wiener ([1948] 1961:121 f.) suggested „a very satisfactory method for constructing a short-time memory is to keep a sequence of impulses traveling around in a closed circuit until this circuit is cleared by intervention from outside." In his Reith Lectures of 1950 Young (1960) suggested to combine the memory cell theory with the circuit theory. He had identified the memory center in the brain of an octopus (cf. Boycott and Young 1955) and yet held to the 'circuit' model, which he explained by comparison with a telegram that travels around between cities so that its message remains intact (Young 1960:34), because in the octopus nerve-fibers carry impulses to and from the optic lobes. „There is here therefore a circuit that could keep going," and „it seems likely that the method of storing involves in some way the setting up of continuous processes […]" (35 f.). Cf. the remark by Jefferson in his last lecture: „Hypothesis about memory: Memory may be in the form of a circuit going around in the brain and brought back through linking together with the circulating thought." RPC box 24, folder 1, p. 2.

[24] Cf. Wiener [1948] 1961:133–143 on scanning.

JEFFERSON spoke of the reduction of the importance of considerations regarding the cerebral cortex in neuro-physiology.[25] Regarding 'petit mal' - in the case of traces taken from cats—we have a 'spike' in the wave-trace which is very similar to what we get from human traces under similar conditions. This is associated with the optic Thalamus.[26]

HEWELL: regarding 'scanning'; the idea is that the networks are being scanned by a discharging system. If a piece of the cortex is removed it continues to discharge in the way that it does before removal.

YOUNG: told of the experiment of cutting into a frog's brain, in which case the pulses resume when the two halves are put back together. Hence 'connexion' does not seem to be essential to the brain-functions as far as the traces show us them. In Octopus there is a centre of cells which, if cut out, result in the animal's being unable to retain the memory of a very simple trick—it retains it now for 5-6 minutes only, whereas the normal retention time is 8-10 days.[27] Hence this seems to be a reinforcement of a cell-theory for memory at least.

NEWMAN: does this support a theory of electrical-charge stora[g]e, like we have in the machine?

..........? said that there was an analogical process in the machine.

HEWELL: this is the difference between the machine and the brain; the brain is 'reminded' by a 'leak' into established circuits: in the machine the current flows around them. (This is McCullough's [sic] memory-store-cells theory[28]).

[25] Cf. Jefferson, „Variations on a Neurological Theme — Cortical Localization" (1955; in Jefferson 1960:35–44), concluces (p. 43): „This is the crux of the new play of integration in which cortical localization has its proper and honoured place. If we no longer believe that such things as love and loyalty, tenderness and friendship and happiness, have special abodes in the brain, yet we know that some such attributes exist, fictitious as definite entities though they may be, and that they are the by-products of the harmony of body and brain."

[26] Cf. report on E.E.G. experiments with cats in Jefferson 1960:526–536.

[27] Cf. Young 1971:621 and Boycott and Young 1955, Young 1960:35.

[28] Warren S. McCulloch (1898-1969). In his collected papers *Embodiments of Mind* (1988) there is no explicit reference to memory-store-cells; however, in „Machines that Think and Want" (1950) he described memory in a similar way as discussed here: „Purposive acts cease when they reach their ends. Only negative feedback so behaves and only it can set the link of a governor to any purpose. By it, we enjoy appetites, which, like records that extend memories, pass out of the body through the world and returning stop the internal eddies that sent them forth. As in reflexes, the goals of appetition are disparate and consequently incommensurable. They may be as incompatible as swallowing and breathing; and we are born with inhibitory links between the arcs of such reflexes. But of appetitions the dominance is rarely innate or complete, and we note the conflict whose outcome we call choice. When two physically or psychologically necessary acts are incompatible, 'God' cannot forgive us for not doing the one because we must do the other. The machine inevitably goes to hell." (1988:310 f.) McCulloch's and Pitt's paper on „A Logical Calculus of the Ideas Immanent in Nervous Activity" of 1943 (1988:19–39) was inspired by Turing's famous paper on „Computable Numbers". McCulloch also visited Turing in Manchester who did not think highly of him. (Hodges1983:304, 343 and 411; Copeland and

YOUNG disagrees on clinical grounds.[29]

JEFFERSON: we have no idea where these 'stores' are, but there is merely no better suggestion. But if part of the brain is removed memory remains.

YOUNG: 'limited space' conception in this matter seems to be false.

NEWMAN: there are some limits, surely, to the possible 'cutting-away' - some 'compression' does take place, but there is after all a limit.

JEFFERSON: yes, but what were the last remaining cells doing in the meantime?

YOUNG: storage seems to be in the whole, and not in any particular part. Gestalts seem to be involved.

NEWMAN: what we ought to do is to start like the atomists with a 'billiard ball' hypothesis - a hypothesis which is obviously wrong, yet which is after all a point of departure.

YOUNG and JEFFERSON agree.

YOUNG: Logic might help to ask the right questions, and to set up hypotheses.

NEUMANN[30]: spoke of attempted consistency proofs as regards the theory of neural networks. (H. Copeland).

NEWMAN: crude models can at any rate be eliminated.

POLANYI: how can e.g. 'seeing stereoscopically' be made the subject of a 'model'? What is the connection?

……? replied that the use lay in guiding advancing hypotheses.

……….?: the question is not one of a 'reality' relation, but of the use to which a model can be put.

NEWMAN: spoke of 'logical similarity' between e.g. animals and the mendelian heredity tables. the [sic] model is to be distinguished from the explanation.

JEFFERSON: said that many of these 'models' are not worth making, because you already know what is going to result from them.

YOUNG agrees logically, but says that 'intuitively' you learn a great deal.

NEWMAN agrees

……….? says that before you get results there must be correspondence between the model and the reality, e.g. neurological model lacks certain correspondence.

Proudfoot 2005:133). The same paper by McCulloch and Pitts might have motivated John von Neumann to work on computers (Nagy et al. 1989:184). A more recent paper by McCulloch and Pitts „How We Know Universals. The Perception of Auditory and Visual Forms" [1947] (McCulloch 1988:46–66) discussed „neural mechanisms … that exhibit recognition of forms".

29 Cf. Young's theory of memory cells in Young 1960, and 1971. Young 1965:297: „The memory system of each animal species consists of a number of modules, provided by heredity (figure 10). Each unit or mnemon consists of a classifying neuron that responds to the occurrence of some particular type of external event that is likely to be relevant to the life of the species."

30 Bernhard Hermann Neumann (1909-2002), who was lecturer in mathematics at the University of Manchester since 1948: Knowles 1959, Debus 1968:1248; Mack and Newman 2003; autobiographical record at http://www.science.org.au/scientists/interviews/n/bn.html.

POLANYI: what meaning have such models? Can we derive from the model the conception of 'seeing in depths'?

NEWMAN: In making models we assume that some quantitative solution is possible, and the rest is left out.

HEWELL: regarding 'choice': implies two or more potential incompatibles, hence there must be an element of choice here, i.e. inhibitory power must be exerted. In the animal a path is established such that the preferred action results.

TURING: yes—random operation can be made to become regular after a certain prevailing tendency has shown itself.

HEWELL: respiratory centres have movements which can not be inhibited, but in choice the incompatible can be accepted and the normal rejected.

TURING: machine may be bed[31] with incompatibles, but when it gets 'contradiction' as a result, there is then a mechanism to go back and look at things which led to the contradiction.

JEFFERSON: but this is an argument against the machine: do human beings do this kind of thing?

TURING: yes—mathematicians.

(Murmur—are mathematicians human beings?)

(Details of this 'going back' process asked for).

NEWMAN suggested that this kind of thing was more on the subject of lines of conduct, and was not covering the logical aspect only.

TURING: declares he will try to get back to the point: he was thinking of the kind of machine which takes problems as objectives, and the rules by which it deals with the problems are different from the objective. Cf. Polanyi's distinction between mechanically following rules about which you know nothing, and rules about which you know.

POLANYI tries to identify[32] rules of the logical system with the rules which determine our own behaviour, and these are quite different things.

EMMET: the vital difference seems to be that a machine is not conscious.[33]

[31] Perhaps: fed. Suggestion from Martin Moleski. The transcription in *Rutherford Journal* and Mays 2000:62 have „bad", but in the copy of RPC the vowel is clearly „e".

[32] Probably to be corrected: „Polanyi: Turing tries to identify rules of the logical system …"

[33] In her Sir Samuel Hall Oration of 1947 Emmet addressed Samuel Alexander's view „that in one sense mind is identical with an organized structure of physiological neural processes. There is no animistic or purely 'mental' factor to be distinguished over and above these. But certain organizations of neural processes have as a function a completely new quality, conscious awareness. In this respect, mind is a new emergent. „ (Emmet 1950:227) Turing (1950:446) will label this as Jefferson's argument from his Lister Oration 1949 (cf. Jefferson 1949:1110) and as „the solipsist view" because consciousness can be known only by the one who is conscious. Ultimately „the only way to know that a *man* thinks is to be that particular man." He then obviates it with the „imitation game" that became known as the Turing Test.

TURING: a machine may act according to two different sets of rules, e.g. if I do an addition sum on the blackboard in two different ways:

(a) by a conscious working towards the solution

(b) by a routine, habitual method

then the operation involves in the first place the particular method by which I perform the addition—this is conscious: and in the second place the neural mechanism is in operation unconsciously all the while. These are two different things, and they should be kept separate.

POLANYI interprets this as suggestion that the semantic function can ultimately be specified; whereas in point of fact a machine is fully specifiable, while a mind is not.

TURING replies that the mind is only said to be unspecifiable because it has not yet been specified; but it is a fact that it would be impossible to find the programme inserted into quite a simple machine - and we are in the same position as regards the brain. The conclusion that the mind is unspecifiable does not follow.[34]

POLANYI says that this should mean that you cannot decide logical problems by empirical methods. The terms by which we specify the operations of the mind are such that they cannot be said to have specified the mind. The specification of the mind implies the presence of unspecified and pro-tanto unspecifiable elements.

TURING feels that this means that my mind as *I know it* cannot be compared to a machine.

POLANYI says that acceptance as a person implies the acceptance of unspecified functions.

..........?: re-raises the point regarding the undiscoverability of [a] programme inserted into machines. Could this be clarified?

Next came a return to the 'model' question as regards memory storage.

YOUNG was unable to see any possible 'picture' of memory storage.

TURING suggested that a machine containing neuron-models might help.

YOUNG gave technical details.

TURING asked what could be taken as model cells?

YOUNG gave suggested diagrams of nerve cells in star-shaped arrangements,[35] and much discussion with TURING ensued.

On this note the meeting closed.

[34] „Turing is obviously here identifying the mind with the brain. Polanyi is not saying that the elements of the brain are unspecifiable but only the mind." Mays 2000:62.

[35] Perhaps diagrams by Golgi/Ramón y Cajal of star-cells as reproduced in McCulloch and Pitts „How We Know Universals" [1947] (McCulloch 1988:55).

2. Can the mind be represented by a machine?[36]

13th September 1949.
Can the mind be represented by a machine?
by Michael Polanyi

Notes for discussion on 27th[37] October 1949.

I take it that the scope of computing machines is identical with that of a formalised logical procedure, i.e. one which handles symbols according to exact rules and also announces results—such as 'A is a proof of B'—by the application of such rules. I shall assume that while no machine is certain to operate faultlessly, we may imagine one which does; and that, while no rule is certain to be unambiguously applicable indefinitely, we may assume this to be true for some formal systems.[38] I suggest that the two propositions are equivalent.

(1) The programme initiated by David Hilbert 50 years ago of reducing mathematics to a specifiable set of axioms operated on according to exact rules has failed. The discoveries of Gödel (1930)[39] have shown that arithmetic and advanced geometry[40] are incomplete; for it has been possible to set up problems of an arithmetical or geometrical character that can be neither positively nor negatively decided in these systems. It was also proved by Gödel that no extension of the axiom system can remedy this deficiency and no consistent and complete deductive theory, containing as its theorems all true formulae of arithmetic and advanced geometry, can ever be constructed.

Finally, Gödel demonstrated that when a formula has been proved undecidable within a given set of axioms we may yet feel compelled to accept its truth and thus extend the pre-existing set of axioms. By reflecting on the very operation by which we proved the undecidability of a formula we are compelled to conclude that it is true and thus add it to the list of our axioms. This process can be extended indefinitely.

There is established thus an inexhaustible procedure for the discovery of ever more true mathematical formulae, which, by its very nature, is incapable of formalisation.[41]

36 RPC box 32, folder 6. Above the header, in the upper left corner, opposite to the date: „84.".

37 Corrected from: 20th. This and a few other corrections are certainly by Polanyi. There are three comments by a different hand (see notes 40, 41 and 44); compared with manuscripts published at The Turing Digital Archive (www.turingarchive.org) they could be by Turing.

38 A question mark in the margin next to this sentence.

39 If the year is intended to be a precise reference Polanyi must be using the abstract Gödel 1930.

40 Superscript by unknown hand: *rather: number theory.*

41 "nature… formalization": underlined and annotated by unknown hand: *no./not in the same language. But we can formalize the meta-language.*

This procedure can therefore be carried out only by a human mind and not by any computing machine.

We have here a function of the mind which cannot be represented by a machine.

(2) I regard this argument as conclusive. As I now proceed to paraphrase and generalise it, any accompanying reduction in precision should not be taken to reflect on it.

The 'informal' (non-formalisable) procedure by which we can generate an indefinite number of new axioms may be described as a process of reflection. For it starts from a contemplation of a mental process of our own and leads thence to new conclusions which lie beyond the previously defined range of our mental operations. This recalls it that according to Poincaré[42] all mathematical innovation is essentially analogous to the procedure of 'mathematical induction', and like this consists in a „bending back of the mind upon itself by which it observes its own mode of reasoning."

These remarks may serve as a transition from Gödel's work on the decision problem to Tarski's on the semantic definition of truth. The original paper on this subject, in Polish, came out shortly after Gödel's. According to Tarski[43] the conclusions previously reached by Gödel can be derived also from his own work, which is in fact based on the process of reasoning essentially similar to that first used by Gödel.

Tarski shows (by working his way backwards from the paradox of the Liar) that no self-consistent formal language can contain any semantic terms—like 'true', 'signifies', etc.—which have application to sentences formed in that language. Thus in no given formalised language can the question be asked whether any statement made in terms of that language is 'true', 'significant' etc. (Though we may say that it is provable or ask whether it is so in that system). In other words: we cannot reflect within any given self-consistent formal language on the truth, significance, etc. of anything expressed in that language.

Our minds however are not similarly limited.[44] We are always capable of reflecting significantly on the truth or meaning of any statement we have made before. Our mental powers are therefore essentially wider than are the operations of any given formal language.

Tarski has shown that these reflective powers of our mind, which no formal system can adequately represent, form an inexhaustible source of innovation. They produce meta-languages that are 'essentially richer' than the language used for the statements on which we are reflecting. We have here once more Gödel's indefinitely extending programme of innovation, which can be achieved only by informal methods and not by a machine.

[42] Polanyi's footnote: Poincaré, „L'intiution et la logique en mathematique <sic>„ (1900)

[43] Polanyi's footnote: Tarski, Journal of Philosophy and Phenomenological Research, Vol. 4, 1944:342–375. [Correct: *Philosophy and Phenomenological Research*, 341–376.]

[44] Superscript by unknown hand: *But they are. Otherwise we get into the paradoxes.*

(3) The great advantage to be derived from the work of Gödel and Tarski is the precision it has lent to a knowledge that we have long possessed in vaguer forms. The argument that a machine differs from a human being in that it cannot reflect upon itself, is common enough. Yet it failed to cut through the tangles of determinism for it did not exhibit a definite representation of the process of reflection and of the innovations which flow from it, as opposed to unreflecting processes of current reasoning.

A parallel distinction that can be drawn between 'rules' and 'interpretation of rules' can be perhaps more successfully established by informal[45] argument. We are constantly applying in many fields of life, rules that cannot be applied according to any exact formal criteria. This has been most forcibly brought out in the field of judicial procedure. There have been repeated attempts (in France and Switzerland) to eliminate the discretion of judges and bind them to a 'strict' application of the law. It was desired to suppress thereby the process of legal interpretation as an informal legislative power.[46]

The attempts failed. Inevitably there grew up around the codes a system of interpretations. No explicit rules can ever suffice to ensure their own rational application to local cases, and the need for interpretation is therefore irresistible.

I shall quote the passage from Kant's Critique of Pure Reason to show how he defines the human faculty of judgment as a residue which no formalisation can exhaust.

> If it were attempted to show in general how anything should be arranged under ..... rules, and how we should determine whether something falls under them or not, this could only take place by means of a new rule. This, because it is a new rule, requires a new precept for the faculty of judgment, and we thus learn that, though the understanding is capable of being improved and instructed by means of rules, the faculty of judgment is a special talent which cannot be taught, but must be practised.[47]

A faculty 'that cannot be taught' can neither be embodied in a machine. Judgment, as here defined, is such a faculty. If we claim to possess it, we affirm by the same token that the working of our minds transcends to this extent any representation by a system of rigorous computations or by machines carrying out such computation.

(4) Judgment, which transcends any system of strict rules, is also indispensable for the hoarding of any beliefs based on experience and for the pursuit of new

[45] informal: ms. correction over: purely logical.

[46] Polanyi's footnote: Comp. J. Walter Jones, „Historical Introduction to the Theory of Law", p. 53. (Oxford, Clarendon Press, 1940).

[47] Immanuel Kant's *Critique of Pure Reason*, transl. by F. Max Müller. New York: Macmillan, 1920, Transcendental Analytic, book II, Introduction, p. 109. Only the beginning is different: „If it [sc. understanding] were to attempt to show..."

discoveries by the process of empirical research. There exists no formal criteria by which we can distinguish between those empirical conclusions which we accept and those that we reject. Our decision between the two is necessarily informal and therefore cannot be represented by the operations of a machine.

We may say that whenever we rely on induction we express by implication a belief in a faculty of the mind which a machine cannot possess. And since in fact we never cease placing our reliance on empirical conclusions, consistency requires that we should deny the possibility of representing the mind by a mechanical model.

(5) The conclusions arrived at in these notes amount to a denial of mechanical determinism. Either the brain works as a machine, then the mind cannot be determined by the functions of the brain; or, if the mind is so determined, then the cerebral processes cannot have a mechanical structure, but must be governed by some other hitherto inconceivable laws which would enable these processes to represent the non-formalisable operations of our mind.

3. Jefferson's last lecture[48]

Jefferson's last lecture. 11,2,1952.
The Working of the Human Mind.

Brain is to be taken at the Morse-code level. Digits correspond to nerve cells.
One can say that machines 'think'; they solve chess problems, but are very slow. The machine takes 15 minutes over a solution where man requires a few seconds. But it is maintained that the question was not put in the best way, that is why the machine took so long over it.

In any case, it was for the first time that the machine had played chess. Same with humans when playing first.

If machine told to discard anything but solution, it did it in no time. This is learning. When machine repeats this performance, then it is remembering. It is maintained that the machine remembers for 75 years. But information (=memory) will have to be refreshed in between. Same holds for humans.

We say that it can be called thinking -of a sort. We make this reservation, as it is feared that humans might be considered mere machines. But it is thinking.

–·–

[48] RPC box 24, folder 1. The text is typewritten. On the fist page, upper right margin is written in ink: JEFFERSON, not Polanyi's hand. However, the date (day-month-year), the typographical errors, and some phrases suggest that the writer was not a native speaker of English. So it is reasonable to assume that Polanyi had typed the notes. The second part is a transcription of the original flyer that invited to the lecture series.

Means of awareness, the fact of self, may never be known. One last final will inevitably elude us. (Whitehead).

–·–

The structure of the brain, as well as the functions of the structure of the brain are all mechanical. Machines are excellent in mathematics, human beings are not. But people will say, that machines are designed for the purpose.

Maths = logic. Man is logical only periodically as he has too many thoughts, is aware of too many contingencies, which block his logic. If logic operates in thought alone, disrigarding [sic] everything else, the man may shock the world. Such as completely logical people like S[c]hopenhauer, etc.

The cleverness of the machine depends on how cleverly the programme was set to it. Same applies to humans: we are being told, taught, we hear and see things and in the end learn everything. Amounts to the same.

–·–

Men were not created equal. Animals are unequal too, as proved .from recent researches. For instance dogs trained and then 40% failed to qualify for police-dogs. We don't know whether animals clever at birth, or when cleverness develops.

Morley mad[e] experiments on ant colonies. Maze problems set and the 'leaders' removed. The whole ant colony stopped, became histerical [sic], etc. When leaders, put back, maze again negotiated. Variations in animals go right down the line of animals.

Thomas Huxley on pugnacity. Man at war with environment, this explains human progress.

–·–

When impulses enter the brain, there are no words attached to them, except when we read. For instances we know people, without remembering their names. Sensory impression come into the brain and we put the words to them. This is how art and literature developed.

[(]Question in discussion: would he say that music developed the same way? Answer: something left out, namely emotion.)

Animals may have some sort of a language, but we don't think so. Their brains not complicated enough.

Hypothesis about memory: Memory may be in the form of a circuit going around in the brain and brought back through linking together with the circulating thought.

Soul: We shan't know until everything the man does, is, etc. will have been explained mechanically. The soul[49] the very last mechanically unexplainable thing which will remain.

49 Corrected from: sould.

Machines may help in solving how man thinks; though up to now they show thinking, as they can be constructed in many different ways there is no reason to assume that in their present form they show how the nervous system operates.

Machine good at arithmetic, bad at maths.

The human mind can evolve something new, the machine can't. Though one is sometimes surprised at the results produced by the machine, the reaction is always that one might have thought of it in advance, it was just overlooked. Nothing basically new, inherent in the construction of the machine.

[Flyer]

University of Manchester
Extra-Mural Department
Session 1951-52
University Extension Course of Three Lectures
The Workings of the Human Mind
by
Professor Sir Geoffrey Jefferson, F.R.S.,
Emeritus Professor of Neuro-Surgery
Tuesdays at 7 p.m. In the University (Room 7, beginning January 29th. Main Building) Fee 2s. 6d.
Tickets for the course *can* be *obtained, on payment of the* fee,
from the Director of *Extra-Mural Studies, The University, Manchester, 13.*

SYLLABUS
LECTURE I.
January 29th.
Nature of nervous system. Essentially a fast communication system. Near and distance receptors. Relation to rest of body. Nerve fibres and nerve cells. Activity of cells and groupings in nuclei. The" nervous Impulse". Main subdivisions of brain and spinal cord. Their uses. The different areas. The" visceral brain". The brain's control of body's chemistry. Nature of emotions. The cerebral cortex. Discovery of its electrical excitability. Localisation of function, phrenology old and new.

LECTURE II.
February 5th.
Nature of thinking. Use of words in thinking. Absence of words in much thinking. Thinking in animals. Man's use of words, failure of stimulation of the brain to produce verbal pictures or speech. Discovery of the speech areas and their history since: Aphasia. 'Dreams are thinking during sleep. Consciousness and sleep, relationships and nervous mechanisms.

LECTURE III.
February 12th.
Dawn of mechanical interpretation of brain and mind. Anatomical searches for the soul from seventeenth century onwards. Mind not an entity but a supposition. Modern views on mechanisms of brain actions in thinking, remembering. The puzzle of differing individual ability to synthesise experience. Difference in animal abilities no less than man's. Imagination. Creative thinking. Art and science similarities of. If man is a machine, what is a machine? How far can proof go? Whitehead's „final irrelevance".

Bibliography

Blum, Paul Richard. 2010a. The Immortality of the Intellect Revived: Michael Polanyi and His Debate with Alan M. Turing. In: *Knowing and Being: Perspectives on the Philosophy of Michael Polanyi*, ed. Tihamér Margitay. 173–184. Newcastle upon Tyne: Cambridge Scholars.

Blum, Paul Richard. 2010b. Michael Polanyi: The Anthropology of Intellectual History. *Studies in East European Thought* Vol. 62:197–216.

Boycott, B. B. and J. Z. Young. 1955. A Memory System in Octopus vulgaris Lamarck. *Proceedings of the Royal Society of London. Series B, Biological Sciences* Vol. 143, No. 913(May 17, 1955):449-480.

Charlton, H. B. 1951. *Portrait of a University, 1851-1951, To Commemorate the Centenary of Manchester University*. Manchester: University Press.

Copeland, B. Jack and Diane Proudfoot. 2005. Turing and the computer. In: (*Alan Turing's Automatic Computing Engine. The Master Codebreaker's Struggle to Build the Modern Computer*, ed. B. Jack Copeland. 107–148. Oxford: University Press.

Debus, Allan G. et al. (eds.) (1968). *World Who's Who in Science*, Chicago: Marquis-Who's Who.

Emmet, Dorothy. 1950. Time Is the Mind of Space. *Philosophy* Vol. 25, No. 94(Jul., 1950):225–234.

Gani, Joe. 2002. Maurice S. Bartlett. *Journal of Applied Probability* Vol. 39, No. 3(Sept. 2002):664–670.

Gödel, Kurt. 1930. Einige metamathematische Resultate über Entscheidungsdefinitheit und Widerspruchsfreiheit, *Anzeiger der Akademie der Wissenschaften in Wien* 67:214-215; reprinted in *Collected Works*, Vol. 1, ed. Solomon Feferman et al. 140-142. New York-Oxford: University Press-Clarendon Press. 1986.; with translation „Some meta-mathematical results on completeness and consistency" (1930b).

Gödel, Kurt (1931). Über formal unentscheidbare Sätze der Principia mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik* 38:173-215; reprinted in *Collected Works*, Vol. 1, ed. Solomon Feferman et al. 144-195. New York-Oxford: University Press-Clarendon Press. 1986.; with translation „On formally undecidable propositions of *Principia mathematica* and related systems I" (1931).

Hagner, Michael. 2008. *Homo cerebralis. Der Wandel vom Seelenorgan zum Gehirn*. Frankfurt: Suhrkamp.

Hodges, Andrew. 1983. *Alan Turing: The Enigma*. New York: Simon and Schuster.

Hodges, Andrew. 1999. *Turing. A Natural Philosopher*. New York: Routledge.

Hofstadter, Douglass S. 1999. *Gödel, Escher, Bach: an Eternal Golden Braid*. New York: Basic Books.

Jefferson, Geoffrey. 1949. The Mind of Mechanical Man. *British Medical Journal* Issue no. 4616 (June 25 1949): 1105–1110.

Jefferson, Geoffrey. 1960. *Selected Papers*. Springfield, Ill.: Thomas.

Knowles, Vincent and University of Manchester. 1959. *Register of Graduates and Holders of Diplomas and Certificates 1851-1958*. Manchester: University Press.

Mack, J. M. and M. F. Newman. 2003. Vale Bernhard Neumann 1909-2002. *Australian Mathematics Teacher* Vol. 59, Issue 1(March 2003):20–22.

Margitay, Tihamér. 2010. From Epistemology to Ontology: Polanyi's Arguments for the Layered Ontology. In: *Knowing and Being: Perspectives on the Philosophy of Michael Polanyi*, ed. Tihamér Margitay. 128–140. Newcastle upon Tyne: Cambridge Scholars.

Mays, Wolfe. 2000. Turing and Polanyi on Minds and Machines. *Appraisal* Vol. 3(October 2000):55–62.

Mazlish, Bruce. 1967. The Fourth Discontinuity. *Technology and Culture* Vol. 8:1–15.

McCulloch, Warren S. [1965] 1988. *Embodiments of Mind*. Cambridge, Mass.: The MIT Press.

Mullins, Phil. 2010. Michael Polanyi's Use of Gestalt Psychology. In: *Knowing and Being: Perspectives on the Philosophy of Michael Polanyi*, ed. Tihamér Margitay. 10–29. Newcastle upon Tyne: Cambridge Scholars.

Nagy, Dénes, Péter Horváth, and Ferenc Nagy. 1989. The von Neumann-Ortvay Connection. *Annals of the History of Computing* 11:183–188.

Penrose, Roger. 1994. *Shadows of the Mind. A Search for the Missing Science of Consciousness*. Oxford: University Press.

Piaget, Jean. 1969. *The Mechanisms of Perception*. New York: Basic Books.

Polanyi, Michael. 1950. Scientific Beliefs. *Ethics* Vol. 61, No. 1(Oct., 1950):27-37.

Polanyi, Michael. 1952. The Hypothesis of Cybernetics. *The British Journal for the Philosophy of Science* Vol. 2, No. 8(Feb., 1952):312-315.

Polanyi, Michael. 1967. *The Tacit Dimension*. Garden City, N.Y.: Anchor Books.

Polanyi, Michael. 1974. *Personal Knowledge. Towards a Post-Critical Philosophy*. Chicago: University Press.

Scott, William Taussig and Martin X. Moleski. 2005. *Michael Polanyi. Scientist and Philosopher*. Oxford: University Press.

Turing, A. M. 1939. .Systems of Logic Based on Ordinals. *Proceedings of the London Mathematical Society* Series 2, Vol. 49:161–228; reprinted in Idem. *Mathematical Logic*, ed. R. O. Gandy and C.E.M. Yates. 71-148. Amsterdam etc.: Elsevier, 2001.

Turing, A. M. 1950. Computing Machinery and Intelligence. *Mind* 59, No. 236:433-460; reprinted in Idem. *Mechanical Intelligence*, ed. D. C. Ince. 133-160. Amsterdam etc.: North-Holland, 1992.

Turing, A. M. 2004. *The Essential Turing*, ed. B. Jack Copeland. Oxford: Clarendon Press.

Von Neumann, John. 1967. The General and Logical Theory of Automata. In *Cerebral Mechanisms in Behavior*, ed. Lloyd A. Jeffress. 1-41. New York and London: Hafner (facsimile of the 1951 edition).

Wiener, Norbert. [1948] 1961. *Cybernetics or control and communication in the animal and the machine.* Cambridge, Mass.: The MIT Press.

Young, J. Z. 1960 *Doubt and Certainty in Science. A Biologist's Reflections on the Brain*. New York: Oxford University Press.

Young, J. Z. 1971 *An Introduction to the Study of Man*. Oxford: Clarendon Press.

Young, J. Z. 1965. The Organization of a Memory System. *Proceedings of the Royal Society of London. Series B, Biological Sciences* Vol. 163, No. 992(Nov. 23, 1965):285-320