Melampus

Could John Lucas be right after all?

The Rationalist Philosophy of Mathematics

The dialectic

John Lucas proposed that "Gödel's theorem shows that mathematical insight need not be algorithmic"[1] hence refuting the mechanist claim that the human mind is a Turing machine. During the exchange of papers that followed, he made a very perspicuous remark stating that "The argument is a dialectical one."[2] A dialectical debate occurs when two opposed views are pitted against each other and where neither side agrees upon the premise of the other.

A mechanist is someone who believes that that the human mind is a computer – more specifically, the human mind is a machine that can be modelled by Turing's analysis of computability. Church's Thesis states that whatever is Turing computable is equivalent to a recursive function and that both together constitute a full analysis of what it is to be an algorithm. The mechanist thesis is equivalent to the claim that all valid mathematical inference can be expressed in the language of recursive functions, which is a theory embedded in first-order logic. As Wolff puts it: ""Most logicians (though perhaps not most mathematicians) are convinced that all correct proofs in mathematics could, with enough effort, be translated into formal proofs of first-order logic."[3] First-order logic is a system of inference that admits quantification only over individuals; in second-order logic there is quantification over properties. Therefore, the mechanist also believes that no second-order inference that cannot be reduced to a first-order inference is meaningful and valid. If there exists a second-order inference that is (a) meaningful, (b) valid, and (c) not reducible to a first-order inference, then that would constitute a refutation of mechanism, since such an inference could not be recursive.

The strong cultural movement in favour of the project of building a computer that will simulate human behaviour in its entirety and thus pass the Turing Test has alarmed the minority that still cling to the beliefs of a bygone age – the freedom of the will, the immateriality of the soul and immortality. Lucas is one such person, a practising Anglican, who seeks in one stroke to halt the progressive advance of mechanism as a cultural norm.

Let us define a "monster" to be a valid argument that could not be formalised within a first-order language. Then Lucas believes that he has found such a monster in Gödel's theorem. Not, however, that the theorem itself could not be so formalised, but that the implications of that theorem cannot.

> Granted that no false formula can be proved in Elementary Number Theory, it follows that the Gödelian formula is both true and unprovable from Peano's axioms. I thought I could apply this to the mechanist hypothesis that the human mind was, or could at least be represented by a Turing machine … there would be a Gödelian formula which could not be proved in the formal system and could be seen to be true by a competent mathematician who understood Gödel's proof.[4]

What demonstrates that this argument is dialectical is that there is general agreement between the two sides on the "facts":

(1) The proof of Gödel’s theorem is first-order.
(2) The statement of the theorem is conditional: If Peano Arithmetic is consistent and the theory is “sufficiently strong”, then Gödel’s theorem is true.
(3) No machine that is John Lucas has been specified.
(4) If a *K* is a first-order theory for which the Gödelian formula *X* is true but not provable, then $K \cup X$ is also a first-order theory in which *X* is both true and provable.

Lucas argues that *in addition to all this* the mind can “see” that if a mind be identified with any given machine, that a contradiction ensues. The emphasis is on “seeing” – that is, a species of mathematical intuition, which the mechanist would never grant. The mechanist replies that the human mind is limited just in the same way any Turing machine is limited[5]. He glosses that human creativity is limited in the same way creativity of computers is limited[6]. But the limitation of the human mind in this way is a premise that the non-mechanist would never grant. Thus, the argument is a dialectical one.

A mechanist must be an empiricist, for to allow for non-empirical knowledge is to grant the non-mechanist the very premise that is in dispute. Therefore, it is fitting to describe the non-mechanist as a rationalist – since this is the viewpoint that is most characteristically opposed to empiricism. Penrose represents a viewpoint that is materialist, empiricist and yet non-mechanist, but for heuristic clarity I shall present the non-mechanist as a rationalist – one who believes (a) that knowledge can be of non-material objects called properties, concepts or universals, and (b) that hence the mind is equipped with a non-material, transcendental faculty that traditionally has been called “reason”. In its fundamental character the dialectical debate between the mechanist and non-mechanist is a manifestation of the age-old dispute between empiricism and rationalism.

> Empiricism: All knowledge is derived from sense-experience.
> Rationalism: There exist concepts that are (a) sources of (infallible) knowledge, and (b) not derived from sense-experience.

Another point about this dialectic: regardless of whether to the mechanist/empiricist the doctrine seems strange or not, the pure rationalist asserts that *there is no material basis to the mind*. How, one may ask, can mental faculties subsist without being grounded in material processes? However, that is not an argument encompassed by the debate; the question is whether the view that the mind is immaterial can be supported by examination of its faculty for knowing by its ability to make inferences.

First-order logic provides a series of algorithmic procedures that recursively generate proofs. Any such proof can be encoded in a binary machine. Hence, first order logic is mechanical. The algorithmic procedures provide a syntax. First-order logic is complete: any statement of first-order logic that is true within-the-system, can be proven mechanically to be so. There exists a decision procedure for truth in first-order logic.

> Mechanism (Formalism): Syntax equals semantics
> Rationalism: Syntax is not equal to semantics; there exist proofs that are not encoded in a syntax.

Gödel’s theorem appears to offer the rationalist a single clear mathematical proof that is not encoded in a syntax. However, since the proof itself is a first-order proof, the rationalist claim must go outside the proof, and appeal to mathematical intuition ("We just see it"), which the empiricist denies; hence the argument is dialectical since the disputed premise is: There exists a mathematical intuition which enables us to "see" that the Gödel's theorem has consequence over and above those of its the first-order proof.

Ways to the resolution of a dialectical debate: (1) finding a premise on which both parties agree; (2) higher resolution as in Kant – a synthesis of thesis and antithesis; (3) conversion based on accumulation of evidence and insight; (4) conversion based on psychological investigation of motives for holding a viewpoint discovered to be held in bad faith.

Kant's *Critique of Pure Reason* was an attempt to reach a higher-level synthesis of empiricism and rationalism, exemplified in the four antinomies. In each case, the apparent conflict between the empiricist thesis and the rationalist anti-thesis is resolved by appeal to the distinction between empirical and transcendental reality; the rationalist claim to have direct access to noumena in transcendental reality is rejected, but the empiricist claim that all knowledge comes from sense data is also shown to be false, since there is knowledge that encodes our ability to understand the world of phenomena, and this is said to be synthetic a priori. Kant's idealism historically formed the framework of C19th philosophy.

In the C20th the Kantian synthesis was overthrown by the empiricists. Landmarks of this cultural development include the influence of the Vienna circle, the essay *Two Dogmas of Empiricism* by Quine in which the notion of the synthetic a priori is rejected, Moore's *Proof of an External World*, the emergence of the American New School of Realism, the rejection of the logic of intensions propounded by Bousanquet, Bradley and Husserl in favour of the logic of extensions, advanced by Russell; the influence of Wittgenstein.

Since the empiricists rejected the Kantian synthesis, they in effect *restored the dialectic of the pre-Kantian period*. Then, two worldwide schools of philosophy might have arisen, one representing the modern update of the C18th empiricism advanced by Hume, and the other the modern update of the rationalism of the C17th found in Descartes or Leibniz. This did not happen. Rationalism was overwhelmed and came scarcely to be represented in academic circles; the empiricists triumphed right across the board; scarcely any quarter was left to the rationalists.

By rejecting the Kantian synthesis, the empiricists also rejected the Kantian solution to the problem of freewill[7], and thus reinstated that dialectic.

Mechanism: that the mind is determined
Rationalism: that the mind (reason) is possessed of freewill

The development of logic lies at the centre of the mechanist/empiricist movement. If logic is the science of inference based on intensions mediated by phenomenological enquiry (a theory of judgement), then the empiricist case collapses immediately. Logic, therefore, came to be treated only as a science of extensions; hence, since only first-order logic deals unambiguously in extensions, that was the only permissible logic.[8]

Regarding the philosophy of mathematics with the collapse of the neo-Kantian solution, there initially emerged: (1) logicism, (2) formalism[9], (3) intuitionism, (4) Hibertism[10]. All four theories have coalesced into formalism, which is the dominant theory of the contemporary period, and the empiricist solution. *No rationalist philosophy of mathematics has been forthcoming*.

Hence, from a socio-cultural point-of-view, no attempt to use a single result (Gödel's theorem) could possibly succeed in over-turning the spirit of the age, which is empiricist and mechanist: for such a program to succeed, one must obtain the consent of empiricists to a premise that is equivalent to denying their empiricism. During the course of the century: (1) first-order theory was developed to an advanced state, and academic circles cast off the mixed, Kantian, phenomenalist logic of the previous period in favour of the mechanical first-order logic; (2) computer science was developed and sophisticated digital machines produced; (3) empiricism overturned neo-Kantian explanations of science, for instance, the instrumentalist Copenhagen interpretation of quantum mechanics, to the extent that it appeared that *every advance of science is an advance for empiricism (and materialism)*; (4) the Christian cosmology collapsed and was replaced by the cosmology of the Big Bang and evolution, both of which were captured by empiricists as endorsements of their world-view.

Hence, the attempt of Lucas was doomed to failure, not because of the intrinsic demerits of the argument per se, but because Lucas was swimming against the tide. A single rationalist argument out of context of a systematic rationalist philosophy without support of rationalists in general, who either ceased to exist or became very thin on the ground, could not succeed in convincing the culturally dominant mechanists, with their mesmerising project of developing artificial intelligence.

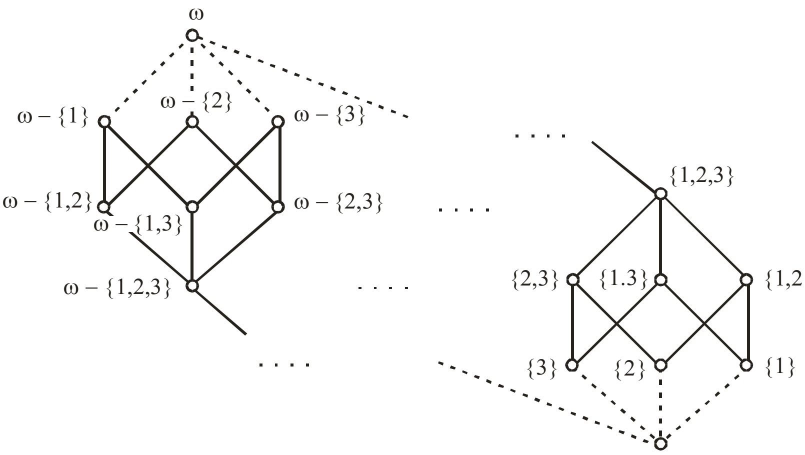
Gödel's theorem

Dialectical arguments can also be resolved by accumulation of evidence, and it is in this spirit that I propose to make observations about the mathematics of Gödel's theorem that may have been overlooked, and in the balance of opinion, these observations could tip the argument in favour of the Lucas point-of-view. That the mathematics I shall adduce could not possibly be first-order mathematics, I will allow; but it may be convincing nonetheless.

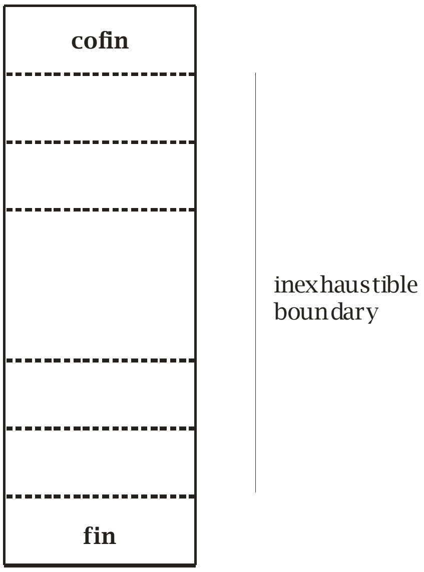
The technical observations I shall make primarily concern models of Gödel's theorem. The continuum is a model in which Gödel's theorem is true, and the Cantor set is a representing set for the continuum. So, I shall begin by making some observations about the Cantor set. The Cantor set is the power set $\omega = \{1,2,3,\ldots\}$, hence comprises all subsets of ω. These subsets are further partitioned; for example, we meet the collections of finite and co-finite subsets:

$$\begin{aligned}&\mathbf{fin} = 2^{<\omega} = \{\varnothing,\{1\},\{2\},\ldots,\{1,2\},\{1,3\},\ldots,\{2,3\},\ldots\{1,2,3\},\ldots\}\\&\mathbf{cofin} \cong 2^{<\omega} = \{\omega-\{1\},\omega-\{2\},\ldots,\omega-\{1,2\},\omega-\{1,3\},\ldots,\omega-\{2,3\},\ldots,\omega-\{1,2,3\},\ldots\}\end{aligned}$$

Between **fin** and **cofin** there is an inexhaustible boundary of sets, belonging to neither.



This boundary is also partitioned into continuum many segments: the boundary contains continuum many collections of sets that can never be enumerated. Each of these segments is created by first removing an infinite subset of ω, and then, from the remaining set adding finite or removing cofinite subsets.



Model of the Cantor set

Let us revert to Gödel's theorem. The Gödel sentence that is at the heart of this theorem is a statement of the form, that for a first-order logic *K*:

$X \equiv$ There is not a proof of *X* $\qquad$ $X \equiv \nvdash X$

The proof of Gödel's Theorem falls into two parts:

> (1) A proof that for a consistent, sufficient strong, first order logic *K*, the Gödel sentence, *X*, is recursive together with the coding, called Gödel numbering, that produces the Gödel sentence.
> (2) A proof for *K* that the Gödel sentence cannot be proven within the given system.

It is the first part that is "difficult"; the second is not so difficult.

To prove, in *K*	$\nvdash X$
Proof by contradiction	
Suppose there exists in *K* a proof of:	$\vdash X$
	$\vdash \nvdash X$
	$\nvdash X$
	$\vdash X$ and $\nvdash X$
Therefore, by contradiction	$\nvdash X$

We can also show that $\nvdash \neg X$, but the argument depends on Gödel numbering, and I omit the technical details here. Together, $\vdash \neg X$ and $\nvdash \neg X$ imply that in *K*, *X* is true; that is:

$\vDash_K X$ but not $\vdash_K X$

Hence, *K* is incomplete: there exists a statement true in *K* but not provable in *K*.

A sufficiently strong theory can express the operations of both addition and multiplication. A theory that can express only addition is not sufficiently strong. Certain theories, for which a version of the Gödel's theorem can be derived, can be shown to be "inessentially" incomplete, in the sense that the incompleteness can be removed. Hence, there is a distinction between essentially incomplete, and essentially complete first-order theories. Additionally, for first-order logic alone, we have a completeness theorem, and indeed, can provide an algorithm for deciding in the positive sense whether a given statement is a theorem of the logic. But for a sufficiently strong first-order theory, we can show that the theory is incomplete, hence no such algorithm exists. Hence, the question:

> What is the difference between an essentially complete and an essentially incomplete first-order theory? How is it possible that first-order logic is complete, whereas a sufficiently strong first-order theory is essentially incomplete?

We consider the models of first-order theories. Clearly, something must have happened to the model of a first-order theory, when sufficiently strong, to make it essentially incomplete. Hence, to answer this question: in an essentially incomplete first-order theory, the model of the theory is constrained in such a way that it must identify as a limit a certain set that I shall call the "boundary"; first-order logic alone is modelled by all infinite Boolean algebras whatsoever, but a sufficiently strong first-order theory must have Boolean algebras that imply the existence as a limit of a boundary. As there is no tool within first-order theory for adding that limit, the existence of the boundary is implied by the theorem, but not proven to exist within the theorem.

The mathematical description of this situation admits of further description. Let K_0 be a first-order, sufficiently strong theory. We shall call the subscript here the *level* of the theory *K*. Let $K_0 \vdash$ denote everything that can be proven within K_0 and $K_0 \vDash$ denote everything that is true within any model of K_0. At level 0 we have a Gödel sentence, X_0 for which

$K_0 \vDash X_0$ but not $K_0 \vdash X_0$

We then form the theory, $K_1 = K_0 \cup X_0$. Let us call K_1 the Gödel extension of K_0. In this theory, $K_1 \vDash X_0$ and $K_1 \vdash X_0$. However, for this theory we also have another Gödel sentence X_1 for which

$K_1 \vDash X_1$ but not $K_1 \vdash X_1$

and so on. This process of generating Gödel extensions by adding the Gödel sentence at level *n* to the theory *n*, has the character of a limit sequence. The very notion of the Gödel sentence

X is not provable at level *n*   X is true at level *n*

creates a weird image. How is it possible that *X* is true at that level, but not provable? But there is another situation like this in mathematics; a convergent sequence on an open set whose limit is a real number, which we add when we add the boundary of that open set, and close it. Likewise, the sequence of theories

$K_0, K_1, K_2, \ldots$

is a limit sequence, whose convergence is the statement *for all* K_n*, there exists an* X_n. Gödel's theorem is a one-step inference: for a sufficiently strong theory K_n at level *n*, there exists a Gödel sentence such that $K_n \vDash X_n$ but not $K_n \vdash X_n$. Let us call this the Gödel statement, $G(n)$. Lucas claims that we "see" that the sequence goes on *forever*; and hence can conclude that *all sufficiently strong first-order theories whatsoever have a Gödel sentence*. Lucas frequently uses the expression "I see that…" when discussing precisely this argument. However, we can put a little more flesh on this "seeing"; for, if he does see, *how does he see?* Answer: by means of a complete, mathematical induction.

$$\begin{array}{l} G(0) \\ \underline{G(k) \rightarrow G(k+1)} \\ \text{For all } n, G(n) \end{array}$$

If we allow this universalisation, and that the statement, for all sufficiently strong theories *K*, there exists a Gödel statement, then such a statement could not be proven by a recursive algorithm. For suppose that it is proven by a recursive algorithm; then there is a sufficiently strong theory K^* such that $K^* \vdash \text{For all } n, G(n)$. But then K^* proves of itself that there is a Gödel sentence such that

$$K^* \vDash X \text{ but not } K^* \nvdash X$$

Since it proves $K^* \vDash X$ it makes truth recursive, and hence contradicts the incompleteness which it asserts. Another way of "seeing" this, is that K^* would have to be one of the theories which fall under its own scope, and then it would have a Gödel sentence that would not be implied by it, and hence would contradict $K^* \vdash \text{For all } n, G(n)$.

Perhaps this is no more than to flesh out in some further detail the argument Lucas makes to "see" that the meaning of Gödel's theorem goes beyond what it states in first order logic. However, it fleshes it out by demonstrating that the inference from *any sufficiently strong first-order theory* to *all sufficiently strong first-order theories* is not an inference in first-order logic, and is mediated by a form of mathematical induction. If that inference is allowed, Lucas's view of the consequences of Gödel's theorem is correct to the person who allows it. Obviously, the mechanist must not allow it. We once again encounter the dialectical point of division between the mechanist and rationalist.

> Mechanist: this sequence of theory extensions cannot go on for ever – at some finite point we reach a level at which we simply for practical or physical reasons cannot continue; and at that level both the human mind and machine are equivalent.
> Rationalist: We can "see" (by mathematical induction or universalisation) that the sequence goes on *forever*; and hence can conclude that *all theories whatsoever have a Gödel sentence*.

The argument goes to the heart of the dialectic. For the rationalist, mathematical induction and number theory are *not* instances of first-order analytical logic, hence *it is possible to prove something in number theory that is not provable in first-order logic*. The formal contradiction that arises from Gödel's theorem, *granted this premise*, is just an instance of this. However, the mechanist insists that number theory is a sub-theory of first-order set-theory, so for him the application in this case must be erroneous.

To add one final remark about Gödel numbering. This is the device by means of which the apparently self-referential statement X $\equiv$ there is no proof of X is constructed as a recursively generated statement within a first order theory. From the mathematical point of view, it is akin to a contraction mapping of the lattice, a remark that I shall now endeavour to explain.

Take a denumerable language and put the symbols of this language in one-to-one correspondence with all the natural numbers as unit sets: $\{1\},\{2\},\{3\},\ldots$ whose union is $\mathbb{N}=\{1,2,3,\ldots\}$. Use the power set axiom to generate all finite sub-sets of N; then each of those sub-sets corresponds to a formula of the language; this may be generated recursively. This corresponds to the lattice **fin** = set of all finite subsets of $\mathbb{N}$. Then Gödel numbering assigns to each one of these formulas a natural number belonging to $\mathbb{N}=\{1,2,3,\ldots\}$. Hence, Gödel number contracts the lattice **Fin** onto the set $\mathbb{N}$, which initially acted as its generating set or skeleton. But now we can regenerate the lattice as the power set of this new skeleton. Hence, the contraction has not in any absolute sense collapsed the entire lattice onto its skeleton, but has simply shunted over one part of the infinite lattice onto another.

Given that the Cantor set is divided by a boundary, then within that boundary there are a whole denumerable sequence of partitions; so, the Cantor set is divided into an infinite sequence of collections of continuum many sets, and an infinite sequence of continuum many co-finite sets. Gödel numbering shunts the sequence of collections of finite sets downwards, and the sequence of co-finite sets "upwards". What replaces these sets that are thus contracted? Answer, sets are pulled over from the inexhaustible boundary.

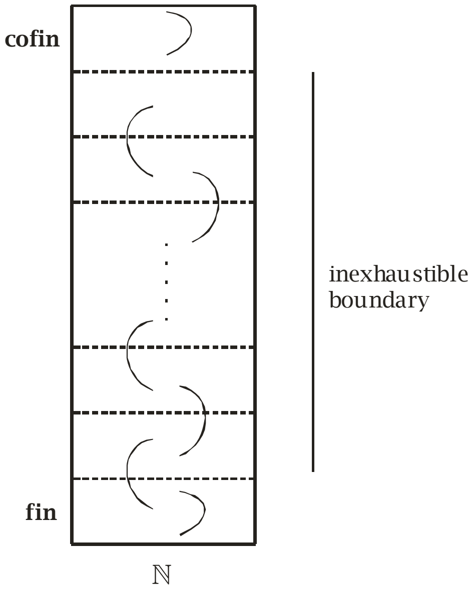


*Diagram illustrating the effect of Gödel numbering on the Cantor set, which is a model of an essentially incomplete theory – to contract **fin** onto the skeleton, and thereafter systematically shunt segments isomorphic to **fin** from and within the inexhaustible boundary*

For a contraction mapping of a Banach space there is always a fixed point – the underlying content of many of the theorems for which Brower was justly famous. The Cantor set is not in this sense a Banach space; nor is Gödel's theorem a fixed-point theorem in that sense. However, the object that plays the role of the fixed point in Gödel numbering is the inexhaustible boundary of the Cantor set.

I believe that should this point be acknowledged, then the discussion of self-reference with regard to Gödel's theorem will be largely perceived to be superfluous; Gödel numbering is not so much a device for encoding self-reference or the Liar paradox, but a contraction mapping of the lattice that is its model.

The existence of this inexhaustible boundary is implied by Gödel's theorem, but is not a first-order consequence of it. Hence, if one allows that the boundary "makes sense", then one is not a mechanist.

The Turing Test

Let us ask: *what would be the effect on the rationalist position of a computer passing the Turing test?* The Turing test arises in the context of the epistemology of other minds – for how do I know that any other given person is conscious? An empiricist will say, by behaviour alone – for there is nothing else for an empiricist to "observe" than behaviour. But a rationalist is not constrained to answer this question in the manner of an empiricist. He might adduce other considerations, even direct intuition. Be that as it may, even for a rationalist the production of a machine that could behave in such a manner that it could fool another person into thinking it was human, would be a considerable advance for the mechanist position. It would put the cat among the pigeons.

It is now nearly seventy years since Turing first made his prediction about computers:

> I believe that in about fifty years' time it will be possible to program computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning. The original question, "Can machines think?" I believe to be too meaningless to deserve discussion. Nevertheless, I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted. I believe that no useful purpose is served by concealing these beliefs.[11]

Yet still we have no machine capable of fooling a living person for more than a few minutes. We receive frequent news on the media of the annual attempt for a machine to breach this barrier, and promises that, with the exponential increase in computing capacity, this vital barrier will be breached in the imminent future. Certain successes are lauded, such as the ability of a computer to out-play Kasparov, break-throughs in playing the game Go, and successes in quiz programs. Algorithms to imitate conversations undergo development.

Will it be possible soon that some of these algorithms may be able to "hold conversations" for some minutes at least with human interlocuters? If so, what would a non-mechanist think? Some things about the Turing Test need to be added.

1. The machine is practising an act of deception. This is akin to any other act of deception. If I go to my Bank Manager and he deceives me into making imprudent investments, then he is a liar; and my being deceived by him, does not make him any less a liar. If a person flatters another person with a view to gaining sexual favours, then he is a liar. Thus, the Turing Test must be placed into the context of deception in general. So, for example, I might hold a

conversation via the telephone with an agent, and be convinced for half an hour, a day, a month, or even years that that agent is a living and conscious human being, only to discover later that *it was a machine after all*. Deception can take place over many years – consider the plot of Henry James's novel, *Portrait of a Lady*. A one-off success in the Turing Test is not as decisive as it may appear to be. To be rigorous, the computer must deceive all the people all the time; even partial failure after the gap of many years will not be sufficient to compel a non-mechanist to reconsider his position.

2. The rigour of the test must also be considered in another respect. Simple conversations will not be sufficient; ordering a cup of coffee in a restaurant according to the "restaurant script" is not a rigorous enough test, for *it is only in the deepest conversations and the profound acts of creation that the human mind is fully expressed*. We must not set the bar too low. In other words, to "pass" the Turing Test, the machine must write a work equivalent to one of Shakespeare, paint a painting like Rembrandt, compose music like Bach and/or produce mathematical theorems and insights like Lebesgue, or equivalent. And this is, after all, the myth that is being propagated by the science-fiction of our contemporary culture – for what do all these stories tell us? – but that, at some time soon, artificial intelligence will surpass that of human intelligence? If that be so, then let the machines compose the novels, the works of art, the music and the science.

And to write a novel, a computer shall have *experience*. It shall have to grow, and learn, encounter other individuals, suffer and so forth: undertake the pilgrimage of life.

The rationalist perspective on the Philosophy of Mathematics that I have sketched here in no way suggests that human intelligence is of such a nature that machines can copy it and exceed it. Human thinking in mathematics does not in its fullest sense even remotely appear to be like that of first-order logic, which is an analytic fragment of all inference whatsoever, and yet the only reasoning that a computer is capable of. A rationalist has no reason at all to suppose that artificial intelligence will be knocking out solutions, for instance, to the continuum problem, or proposing new fixed point theorems, or advancing new definitions of the integral – it is simply of an order beyond them.

So, what of the successes? Well, I observe, for instance, that those successes are the same successes that have been touted for decades. The feature of those *limited* occasions wherein the computer out-performs the human, are that those systems are finite. Chess is a finite problem; so too is the playing of Go; if the knowledge delimited in a quiz program is finite, then an algorithm may beat a human contestant. What of it? Nothing. Human reasoning is most expressed when dealing with the infinite and in creativity.

On this basis, baring the few occasions on which a computer *deceives* an individual for some small space of time, the rationalist *has no reason for believing a computer will pass the Turing Test,* not once the full rigour of the test is instantiated.

What then, of the belief so strongly and widely held that the victory of artificial intelligence is just around the corner? What of the claim, for instance, that the whole universe possibly a computer simulation? This stands at the very extreme of metaphysical speculation of the new religion of artificial intelligence, and yet has not been scorned as spurious speculation, rather met with appreciation and support.

Logically, the statement, "A machine will pass the Turing Test" is in the same class as the claim, "The world will end soon." This latter was a popular belief among the early Christians, and was held by St. Paul for instance. Shall we say that there is evidence for the claims about the Turing Test? Well, there was evidence supporting the claims of the early Christians too, for the Roman world was imbued with moral corruption, so far as they were concerned, and they believed there were signs and portents too.

The evidence for the forthcoming success of the Turing Test is based on the extrapolation of the capacities of computers. Extrapolations are notoriously fallible. Exponential growth has been known to hit a ceiling. In the case of artificial intelligence, the rationalist perspective on the capacities of the mind suggest to him at least that there is an insuperable barrier between the functions of a machine and those of human intelligence – a barrier that no increase in the capacity of a computer could possibly breach – the extrapolation is not justified.

> Mechanist: A machine will be built that will pass the Turing Test even the rigorous one.
> Rationalist: No such machine will ever be built.

Both are statements about the future, while grounded in reasons acknowledged as compelling by each side respectively, none of these reasons are sufficient to compel the opponent. It is not evidence for artificial intelligence now, to say that artificial intelligence in the future shall be better.

To the mechanist: produce your machine, and then let us discuss the consequences. Turing's prediction has not been fulfilled.

An intermediate position?

Is the position of Penrose, who is *both* a materialist and opposed to mechanism, tenable? That one could be an empiricist and non-mechanist seems to me to be entirely possible. This is because a non-materialist philosophy of science is entirely tenable; it may be grounded in Kant, and is the view expressed in the instrumentalism of the Copenhagen interpretation of quantum mechanics. Such a view argues that empirical science identifies the regularities that exist within phenomenal in empirical reality; but eschews the realist assumption that these regularities may be projected onto a transcendent reality of matter; hence, for the instrumentalist, there are only the regularities.

Rationalism obviously does not preclude either an empirical science of the mind, and parts of this science may involve mechanical models of the psyche, cybernetics and such like. For the rationalist, this would be the empirical reflection of the mind, on a principle akin to the relation between the Ego and Self in Kant's philosophy.

While mechanists and techno-realists have assumed that every advance of science is an advance for their philosophy, for the rationalist, this is not so. For the rationalist, science is neutral as to metaphysics, and every empirical law that can be said to cohere with mechanistic materialism, can also be made to cohere with rationalism.

The question for Penrose is what role he assigns to the mind and consciousness in his philosophy. Is the mind the equal partner in the enterprise of understanding phenomena, or is it the product of material forces? As a materialist, he would seem to affirm the latter. As a physicist, he still subscribes to the principle of Galileo that mathematics is the language of reality. He seeks equations that will describe all phenomena, inclusive of the phenomena of the mind. If those equations are written in first-order language and use material implication, then his theory is a mechanical one. The problem arises when his equations are *not written in that language*, for he seems to be committed to some such view, when he advocates a variant of the Gödel argument. If the language he uses is second-order, then, with due respect to the Quinian "to be is to be the value of a variable", he ascribes independent existence to concepts (Platonic realism), and hence brings the mind into equal partnership with matter. Such a theory is not materialist; it is dualist as to explanation.

The future?

Our eyes turn towards the future, and nowhere is the matter of faith more expressed than in our anticipations of the future. If it be acknowledged that the rationalist has an internally viable philosophy of mathematics – then I suggest that we have two faiths in antagonistic view of each other. These faiths conflict in their predictions of the future.

As one sympathetic to rationalism, I must acknowledge that should a machine be built that passes a rigorous Turing test, my understanding of human nature must be shaken, a little at least. It is not shaken now, for I have as many reasons for believing no such machine will be built, as the mechanist can adduce for believing it will.

We await the future. In the meantime, can we live together? It seems to me that in this situation, while both parties await the dictates of fortune, we can and should live together in mutual respect and tolerance. A man who believes, contrary to the mechanist, that such a machine will not be built, has as much right to breathe the free air as another. Freedom of religion, and freedom of conscience are the cornerstones of any liberal society.

Philosophy is the pre-eminent rationalising activity. From Chaucer to Shakespeare to Browning, we have been taught that man builds a system of beliefs that he forces to cohere together, to justify to himself those beliefs that he has stumbled upon within the course of his life, including his self-deceptions and justifications for what he judges in his own conscience to be his moral and immoral acts. At some time in our childhood or adolescence we come across our philosophical views, *readymade*, or almost – for they could not be systematic as yet. One person finds himself a materialist, another a rationalist; but these views are not built upon the foundation of a first-enquiry, as Descartes would define it. Some among us subject our views to as much rigour as our intellectual abilities and personal integrity will allow, for it is a psychological process as much as a logical one; we must plumb the depths of our psyche to uncover not just the grounds of our beliefs in the epistemological sense, but also their motivations. Confronted with such a task fit for heroes, which of us can say in all honesty *that we could not be wrong?* Armed with a humility that comes from thorough self-knowledge, and while anticipating the future, we live, believe and worship in a free society of peers regardless of our differences, no matter how substantial.

References

Benacerraf, Paul [1967]: God, the Devil and Gödel, Monist (1967) 51 (1): 9-32.

Benacerraf, Paul and Putnam, Hilary, Eds. [1987], Philosophy of Mathematics, Selected readings. Second edition. Cambridge University Press, Cambridge.

Bishop, Errett [1967], Foundations of Constructive Analysis, McGraw-Hill, New York

Burton, David M [1976], Elementary Number Theory. Allyn and Bacon Inc. Boston.

Copeland, Jack [2008], The Mathematical Objection: Turing, Gödel, and Penrose on the Mind, July 2008

Curry, Haskell B [1954], Remarks on the definition and nature of mathematics. Original publication in

Henkin, Leon [1971], Monk, J. Donald, Tarski, Alfred: Cylindric Algebras – I, North-Holland

Hamming, R. W. [1998], Mathematics on a Distant Planet, American Mathematical Monthly. Vol. 105. No.7

Lucas, J. R. [1998], The Implications of Gödel’s Theorem, Talk given to the Sigma Club

Lucas, J. R. [1968], Satan Stultified: A Rejoinder to Paul Benacerraf, Monist (1968) 52 (1): 145-158, Dialectica, 8, 1954, 228 – 33. Reprinted in Benacerraf and Putnam [1987] p 202 – 206.

Poincaré, Henri [1996], Science and Method. Trans. Andrew Pyle. Routledge, London 1996.

Potter, Michael [2004], Set Theory and its Philosophy – A Critical Introduction. Oxford University Press.

Turing, Alan [1950], Computing Machinery and Intelligence. Mind LXI (236), p. 433 – 460. Reprinted in Anderson [1964]

Wolf, Robert S [2005], A Tour through Mathematical Logic. The Mathematical Association of America.

[1] The Implications of Gödel’s Theorem

[2] Satan Stultified: A Rejoinder to Paul Benacerraf

[3] Wolf [2005] p.29

[4] The Gödelian Argument

[5] For example, “… what is it that Gödel I precludes the machine (let’s call her Maud) from doing? Evidently, it is to prove H (her Gödel formula) from her axioms according to her rules. But can Lucas do that? Just as evidently not.” – Paul Benacerraf, *God, the Devil and Gödel*.

[6] For example, “The learning mind successively mutates from one theorem-proving Turing machine into another.” Jack Copeland, *The Mathematical Objection: Turing, Gödel, and Penrose on the Mind.*

[7] Third antinomy [Check] – that determinism applies to phenomena that are experienced in time; whereas freewill is a property of the transcendental self, which is timeless, stands outside the time order, and thereby not a product of any successive event taking place in the time order.

[8] The “voice of God” as one acquaintance of mine has put it, though I always thought he was an atheist.

[9] “According to formalism the central concept in mathematics is that of a formal system. Such a system is defined by a set of conventions ... we start with a list of elementary propositions, called axioms, which are true by definition, and then give rules of procedure by means of which further elementary theorems are derived. The proof of an elementary proposition then consists simply in showing that it satisfies the recursive definition of elementary theorem.” (Curry [1954], p. 203)

[10] I distinguish Hilbert’s program from that of the formalism represented by Curry. Hilbert sought finite consistency proofs of ideal postulates in mathematics, but he sought then in order to ground mathematics synthetically; hence, he was not a formalist in the sense of Curry.

[11] From Turing [1950].