THE MECHANICAL TURK

Enduring Misapprehensions Concerning A.I.

A curtain is withdrawn and a machine rolled to within twelve feet of the nearest spectators: a large box of maple wood on castors or brazen rollers with a puppet like a pantomime Turk at the back. His right arm extended to the checker-board on the table, his left supporting a pipe, the Turk is wrapped in a heavy green cloak. But is discovered to consist of mechanical parts, operated by a larger clockwork engine in the box.[1] Having inspected the compartments containing this mechanism, spectators engage the Mechanical Turk in a game of chess: the Turk moving the pieces with his left hand, eyes and head rolling in triumph, on putting his enemy into check he will cry *Échec! Échec!* Built by Wolfgang von Kempelen for the Empress Maria Theresa in 1769, the Turk is said to have won games against Benjamin Franklin and Napoleon Bonaparte, mystifying audiences in Europe and America with a winning-streak lasting 84 years. (Only ended by a fire in 1854, in which the mechanism is said to have perished screaming: *Échec!*).[2] In his own eyewitness account of the Automaton, produced shortly after seeing the chess-player in action while on tour in America in 1836, Edgar Allen Poe suggests that no exhibition of the kind has ever elicited such speculation. 'Wherever seen it has been an object of intense curiosity, to all persons who think', remarks Poe. 'Yet the question of its *modus operandi* is still undecided … accordingly we find every where men of mechanical genius, of great general acuteness, and discriminative understanding, who make no scruple in pronouncing the Automaton a *pure machine,* unconnected with human agency in its movements, and consequently, beyond all comparison, the most astonishing of the inventions of mankind.'[3]

Long after its destruction Von Kempelen's Turk continues to fire the imagination; an eighteenth century fantasy that anticipates the twenty-first century reality we now inhabit; a tangible example of how one era might be said to dream the next. In work by cyberneticians Claude Shannon and Herbert Wiener the Mechanical Turk is represented (in a rhetorical flourish) as precursor to the computer, and thus as harbinger to the processing power that is to bring about the "absolute catastrophe" and "the End of Man". At its peak in the sixties, the "catastrophist" discourse of theorists such as Maurice Blanchot and Lewis Mumford predicted that the coming technological utopia must bring about a perpetual stasis, the end of mankind and historical progress, as every potential future was mapped out in advance by an ever greater processing power. 'The full-blown, the absolute catastrophe would be a true omnipresence of all networks, a total transparency of all data', declares Baudrillard, '… the culminating point of the development of information and communications, which is to say, death.'[4]) Of course, Shannon and Wiener are under no illusions concerning the true nature of the Mechanical Turk. As Walter Benjamin points out, in his *Theses on the Philosophy of History* (1940), the Turkish sorcerer was a confidence trick: operated by a little hunchback sat inside the table, concealed by a system of mirrors, and guiding the puppet's hand by means of strings.[5] But I would like to suggest such playful references to Von Kempelen's fraud in the mid-century literature relating to cybernetics are telling nevertheless. For the computational technologies prefigured by the Mechanical Turk have in fact generated serious and enduring misapprehensions concerning Artificial Intelligence: in the Social Democratic consensus from 1950-1975, and then, subsequently, in a paradigm which is, ostensibly at least, Post-Modern and Neo-Liberal. Although there be too little space here to retrace the theoretical and practical developments whereby computers have impacted upon our perception of history to produce orthodoxies of the End in the postwar period, the following essay will at least discover how that *system of mirrors* has worked; will establish that the threat computers are often represented as posing to human agency and even identity, in material ranging from popular science-fiction to post-humanist philosophy, is in truth, as Poe would seem to be suggesting, nothing less than an uncanny effect.

I.
THE ANALYTICAL ENGINE

In a rigorous piece of investigative journalism, anticipating the methodology and philosophy of his fictional detective M. Dupin, Edgar Allen Poe sets out to reveal the secret of the Mechanical Turk, and concludes it is quite certain the operation of the Automaton is regulated by a human agent and nothing else, going so far as to assert that the matter is susceptible of a mathematical demonstration, *a priori*. 'Let us place the first move in a game of chess, in juxtaposition with the data of an algebraical question, and their great difference will be immediately perceived'.[6] Recalling the 'calculating machine' developed by Charles Babbage, Poe observes that the arithmetical or algebraical calculations performed by this 'engine of wood and metal', however complex, proceed necessarily and inevitably to the one solution that must follow from the data: 'The second step having been a consequence of the data, the third step is equally a consequence of the second, the fourth of the third, the fifth of the fourth, and so on, and not possibly otherwise, to the end'.[7] But from the first move in a game of chess no step necessarily follows; the uncertainty of each ensuing move increasing, one cannot predict with any accuracy more than a few moves ahead. 'Even granting (what should not be granted) that the movements of the Automaton Chess-Player were in themselves determinate, they would be necessarily interrupted and disarranged by the indeterminate will of his antagonist.'[8] So Poe concludes that: 'There is then no analogy whatever between the operations of the Chess-Player, and those of the calculating machine of Mr. Babbage, and if we choose to call the former a pure machine we must be prepared to admit that it is, beyond all comparison, the most wonderful of the inventions of mankind.'[9]

Poe could not have known that the concept of the "calculating machine" underpinning his "demonstration" was being rendered obsolete, even as he wrote, by the inception of Babbage's Analytical Engine. In a series of notes that contain the most important information relating to this project published seven years laters in 1843, Ada Lovelace observes there is 'considerable vagueness and inaccuracy in the minds of persons in general regarding the subject' of Babbage's work. 'There is a misty notion amongst most of those who have attended at all to it,' she writes, 'that *two* "calculating machines" have been successively invented by the same person within the last few years; while others again have never heard but of the one original "calculating machine", and are not aware of there being any extension upon this.' While the first engine developed by Babbage, was strictly *arithmetical*, the results it could arrive at lying within a clearly defined and restricted range, the powers of the new Analytical Engine were co-extensive with our knowledge of the laws of *analysis* itself. 'Indeed', remarks Lovelace, 'we may consider the engine as the *material and mechanical representative of analysis* … the executive manipulation of algebraical and numerical signals.' Inspired by the chains of punched cards employed in mechanised Jacquard looms to regulate the production of complex textile patterns, Babbage had invented a physical system for encoding and uploading a computer program, which anticipated the perforated paper tapes that would later be used to transfer data to the ROM and EPROM on the first minicomputers. 'The bounds of *arithmetic* were … overstepped the moment the idea of applying the cards had emerged; and the Analytical Machine does not occupy common ground with mere "calculating machines",' concludes Lovelace. 'We may say most aptly, that the Analytical Engine *weaves algebraical patterns* just as the Jacquard-loom weaves flowers and leaves.' Indeed, Alan Turing later acknowledged, the machine is at least potentially an example of the "Universal Logic Machine" he laid the groundwork for in his paper 'On Computable Numbers, with an Application to the Entscheidungsproblem' (1936): and *could* therefore have run a chess-program (like the program Turing developed while at Bletchley Park during the Second World War) which assigns numerical values to pieces, positions and potential for future positioning then selects the one move with the greatest value and the greatest position-play value.[10]

One cannot but wonder whether Poe might have been the first person to have realised this. Certainly, there is a marked change of emphasis in his next piece of writing about chess, in the preamble to 'The Murders in the Rue Morgue' (1841), which might suggest that Poe had learned of the Analytical Engine in the interim.

(Poe could not have read Lovelace's notes, nor Babbage's unpublished description of the Analytical Engine in his essay 'On the mathematical powers of the calculating engine' [1937], nor the original Italian-language version of the 'Sketch of the Analytical Engine' produced by L.F. Menabrea (1842), before writing the first of his three stories about Dupin; but Poe was clearly following Babbage's work with considerable interest, and his story might be a response to information concerning the second engine contained in a letter from Babbage to Quetelet published in *Bulletins de l'Académie Royal des Sciences et Belles-Lettres de Bruxelles* (May 1835) and in a collection of essays called *The Ninth Bridgewater Treatise* (1837).) On the one hand, chess is no longer thought to be beyond those 'arithmetical or algebraical calculations' one might perform on a machine.[11] In fact, 'to observe attentively', 'to remember distinctly', 'to have a retentive memory' and 'to proceed by the book', are now considered to be the 'sum total of good playing'.[12] But in both 'Murders in the Rue Morgue' and 'The Purloined Letter', Poe takes issue with popular errors relating to the term *analysis*, promulgated by certain mathematicians.[13] 'The faculty of re-solution is possibly much invigorated by mathematical study, and especially by that highest branch of it which, unjustly, and merely on account of its retrograde operations, has been called, as if *par excellence*, analysis.'[14] But to calculate is *not* to analyse. In fact, the 'constructive or combining power' is a 'primitive faculty', so 'frequently seen in those whose intellect bordered otherwise upon idiocy … as to have attracted general observation among writers on morals'.[15] If this is not generally understood, the 'elaborate frivolity of chess' might help to explain why: 'the pieces have different and bizarre motions, with various and variable values, [and in consequence] what is only complex is mistaken (a not unusual error) for what is profound'.[16] Having rejected 'this particular deception', Poe posits a rather more comprehensive conception of analysis that extends beyond mathematical reason, and the 'pagan fables' of algebra.[17] 'Our player confines himself not at all; nor, because the game is the object, does he reject deductions from things external to the game'.[18] Instead, 'the analyst throws himself into the spirit of his opponent, identifies himself therewith, and not infrequently sees thus, at a glance, the sole methods) sometimes indeed absurdly simple ones) by which he may seduce into error or hurry into miscalculation'.[19] According to Poe, simple games like draughts or whist might, therefore, provide a better means of testing the highest powers of the intellect; in the former, for instance, where the moves are homogenous, and far less calculating power is required, advantages are said to be obtained 'by either party only through superior *acumen*'.[20] In the stories that follow, the analytical powers of Dupin would invariably surpass the immense processing or computational power of the Parisian police-force. — 'His results, brought about by the very soul and essence of method, have in truth, the whole air of intuition.'[21]

Poe's thoughts on chess and analysis were to prove remarkably prescient, anticipating the problems that would preoccupy the mid-twentieth-century pioneers of Artificial Intelligence. As Poe perceived, chess does in fact require far greater computational power than draughts. 'Not only is the end game different from the middle game in the considerations which are paramount, but the openings are much more devoted to getting the pieces into a position of free mobility, for attack and defines that is the middle game', explains Norbert Wiener in *Cybernetics*. 'The result is that we cannot be even approximately content with a uniform evaluation of the various weighting factors for the game as a whole'.[22] And while it would prove easy enough to program machines that could play a 'tolerable game' (Turing) / 'a chess not so manifestly bad as to be ridiculous' (Wiener), the shortcomings of what Poe termed 'an exaggeration of *the application* of the one principle or set of principles of search' soon became all too evident.[23] 'You will find that when the same situation comes up twice on the chessboard, your opponent's reaction will be the same each time, and you will find that he has a very rigid personality', says Wiener. 'It is thus not too hard for an expert to get a line on his machine opponent and to defeat him every time.'[24] What was required clearly was a computer which would perform the sort of observations that inform Poe's Analyst. 'Let us suppose [the machine] examines all the previous games which it has recorded on its memory to determine what weighting of the different evaluations of the worth of pieces, command, mobility, and the like, will conduce most to winning,' suggests Wiener. 'In this way, it learns not only from its own failures but its opponent's successes.'[25] Unfortunately, 'All this is very difficult to do in chess, and as a matter of fact the full development of this technique, so as to give rise to a machine that can play master chess, has not been accomplished.'[26] And that remains the case today, over fifty years later — as philosopher and Turing specialist B. Jack Copeland has pointed out: 'The huge improvement in computer chess since Turing's day owes much more to advances in hardware engineering than to advances in AI.'[27] The victory of IBM's chess-playing computer *Deep Blue* over Russian Grand Master Garry Kasparov was the result not of a learning algorithm but of 256 parallel processors examining 200 million possible moves per second and strategies extending fourteen moves ahead. As Noam Chomsky remarked, the outcome was as meaningful as a bulldozer winning a weight-lifting competition.[28] 'In fact, little or nothing about human thought processes appears to have been learned from the series of projects that culminated in Deep Blue.'[29]

Though this is *not* the impression people formed at the time. 'As Kasparov suspected, his duel with Deep Blue indeed became an icon in musings on the meaning and dignity of human life', observed Robert Wright in *Time Magazine* (1996); 'he seemed to personify some kind of identity crisis that computers have induced in our species.'[30] In fact, the popular reaction to the triumph of IBM is consistent with a tendency to invest chess-playing computers with deeply-rooted anxieties concerning the development of Artificial Intelligence. The first robot in English literature was 'Moxon's Master' (1893), a chess-playing automaton inspired by Von Kempelen's Turk: 'not more than five feet in height, with proportions suggesting those of a gorilla — a tremendous breadth of shoulders, thick, short neck and broad, squat head, which had a tangled growth of hair and was topped with a crimson fez.'[31] Beginning with the question that preoccupied *Time Magazine* (does a machine think about the work it is doing?), Ambrose Bierce describes how Moxon is murdered one night after beating his creature at its own game: 'the whole room blazed with a blinding white light that burned into my brain and heart and memory a vivid picture of the combatants on the floor, Moxon underneath, his throat still in the clutch of those iron hands, his head thrown backward, his eyes protruding, his mouth wide open and his tongue thrust out; and — horrible contrast! — upon the painted face of his assassin an expression of tranquil and profound thought, as in the solution of a problem in chess!'[32] The HAL

9000 adopts a similar course of action immediately after playing a game of chess in Stanley Kubrick's *2001: A Space Odyssey* (1968), and the US intercontinental ballistic missile system "Joshua" suggests that Matthew Broderick might prefer a "nice game of chess", before settling for Global Thermonuclear War in John Banham's *War Games* (1983). So, while Poe is surely correct to suggest that what is *merely complex* has been consistently mistaken for what is profound, one cannot but wonder whether this might not itself be a significant mistake, integral to the pervasive cultural unease, which has so long surrounded the rise of the machines.

In his famous essay on 'The Uncanny' (1919), Freud refers to the impression made on us by waxwork figures, ingeniously constructed dolls and automata, and observes that particularly favourable conditions for generating feelings of the uncanny seem to exist if intellectual uncertainty is aroused as to whether a thing is animate or inanimate.[33] Freud recalls that, young children make no sharp distinction between the animate and the inanimate, but are especially fond of treating their dolls as if they were alive, and speculates that the sense of the uncanny might 'derive not from an infantile fear, but from an infantile wish'.[34] — Our inclination to attribute *thought* to automata might have begun as an insurance against extinction, an attempt to cheat death by investing physical objects with *ego*, on the misguided conviction that *thoughts are omnipotent*. 'But these ideas arose on the soil of boundless self-love, the primordial narcissism that dominates the mental life of both the child and primitive man, and when this phase is surmounted, the meaning of the "double" changes: having once been an assurance of immortality, it becomes the uncanny harbinger of death.'[35] The cultural anxieties generated by Artificial Intelligence, together with our consistently overrating the significance of its achievements, must indicate that what we are dealing with here is an uncanny effect. In this one respect, though they possess no technical similarities, IBM's Deep Blue was no less a conjuring-trick than Von Kempelen's Mechanical Turk.

If all this seems inconsequential, note that Turing first conceived of the 'Imitation Game' which has become the foundation for post-humanist philosophies relating to AI *as a game of chess*. Explaining it is not difficult to devise a program that will play a 'not very bad game of chess', Turing proposes a 'little experiment'. — Get three men as subjects for the experiment: A, B, C. Two of these, let us say, A and C are to be rather poor chess players. B is to do no more than work the machine. Two rooms are used with some arrangement for communicating moves, and a game of chess is to be played between C and either A or the machine B. The point of the experiment is to demonstrate that 'C may find it quite difficult to tell which he is playing'.[36] (According to Turing, this was an idealised form of an experiment he had actually done — probably with the Champernownes at Bletchley Park).[37] As we shall see, this is remarkably close to the final form of the so-called 'Turing Test' or 'Imitation Game' that would soon follow — and this must have profound implications for how we interpret an experiment widely regarded as an empirical measure for AI. For this supposedly behaviourist or operationalist criterion for thinking is clearly designed to gauge an uncanny effect. — 'Playing against such a machine', writes Turing, 'gives a definite feeling that one is pitting one's wits against something alive.'[38]

II.
THE IMITATION GAME

In his famous essay 'Computing Machinery and Intelligence' (1950), Turing sets out to investigate whether machines can think. But (for the very reason Descartes refused to consider whether he is a Rational Animal) Turing proposes to replace this question by another, which is closely related to it, but which will be expressed, claims Turing, in relatively unambiguous words. — 'This new form of the problem can be described in terms of a game which we call the "imitation game"'.[39] This party-game is played by three people, a man (A), a woman (B), and an interrogator (C) in a room apart from the other two who may be of either sex. The object of the game for the third player is to determine which is the man and which is the woman; the interrogator will present the other players with a series of questions, and the answers will be mediated in some fashion to ensure the interrogator decides solely upon the answers provided. The object of the game for the first player is to deceive the interrogator into believing he is the woman, while the object for the second player is to prevent this from happening. 'She can add such things as "I am the woman, don't listen to him!" to her answers, but it will avail nothing as the man can make similar remarks.'[40] Having set out the rules of the game Turing then invites us to imagine what will happen when a machine takes the part of the man? This is the 'more accurate form of the question' that replaces the original problem: — "Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman?'[41] — And Turing is in no doubt as to what the answer might be: predicting that machines certainly *could* triumph in such a contest, he even speculates that this might happen by the end of the 20th century. 'I believe that in about fifty years' time it will be possible to programme 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.'[42] And though the original question "Can machines think?" is dismissed as 'too meaningless to deserve discussion', Turing nevertheless affirms his belief that 'by 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.'[43]

Debate has raged ever since as to what Turing might have intended, and the critical literature on the Imitation Game comprises a vast and bewildering array of differing interpretations. The matter is complicated by the fact that in the same essay Turing presents a different version of the game, where the computer plays the part of A and the part of B is 'taken by a man'.[44] This has provoked commentators such as S.G. Sterrett, in 'Turing's Two Tests for Intelligence' (2000), to distinguish between what he terms 1) the "Original Imitation Game", in which the objective is for the computer to impersonate the woman, and 2) the "Standard Interpretation" of the "Turing Test", in which the objective is to imitate a 'man', with the latter generally being understood to mean *human* rather than merely *male*.[45] And though commentators often miss or reject the possibility of such a distinction, there is a general agreement that Turing had really been trying to assess a computer programme's ability to imitate a human being, rather than to simulate a woman. This "Standard Interpretation" might seem to find support in Turing's later articles and broadcasts, where he says simply that 'The idea of the test is that the machine has to try and pretend to be a man', and that our main problem is to 'programme a machine to imitate a brain'.[46] And it is the "Standard Interpretation" that has provided the criterion for the Loebner Prize, a famous annual competition in which computer programmes compete for a gold medal and an award of $1000,000. The outcome of such competitions arouse interest far beyond the small community of specialists in AI because it is widely believed that Turing was offering his test as a means of defining the nebulous term 'thinking'. 'An especially influential behaviourist definition of intelligence was put forward by Turing', writes psychiatrist Ned Block; and this assumption is shared by philosopher John Searle and physicist Roger Penrose in their own landmark assessments. 'The operationalist would say that the computer *thinks* provided that it acts indistinguishably from the way that a person acts when thinking,' writes Penrose 'This viewpoint was argued of very forcefully in a famous article by Alan Turing…'[47] — Though, as Turing's editor, B. Jack Copeland remarks, Turing says quite explicitly that his aim is *not* to give a *definition* of thinking but rather a '*criterion* for "thinking"'.[48] The result of this emphasis on the "Standard Interpretation" of Turing's Test as a means of verifying the claims of what Searle and Penrose term 'strong AI' is that the "Original Imitation Game" is often considered irrelevant or misleading: dismissed by Turing's biographer Andrew Hodges as a 'red herring', and regarded by Turing's biographer as just part of the protocol for scoring the test: 'Will interrogators decide wrongly as often in man-imitates-woman games as they do in computer-imitates-human games?'[49]

In recent years this critical consensus has come under attack following renewed interest in what Sterret terms the "Original Imitation Game". In their recent reception history Ayse Pinar Saygin, Ilyas Cicekli and Varol Akman note that, while the woman vanishes altogether in the second version of the game which underpins the "Standard Interpretation", the objectives of A, B and the Interrogator remain unaltered, as Turing never explicitly says these have changed; that both the machine and the man are still impersonating a woman.[50] In their opinion, 'The man and the machine are measured in terms of their respective performance against real women.'[51] And they suggest that, though there is an ambiguity in the paper, the change in the second game is intended to stress the point that the simulations are to be compared against each other and not that which they are simulating. 'On close examination it can be seen that what Turing proposes is to compare the machine's success against that of the man, *not* to look at whether it "beats" the woman in the IG.'[52] Saygin, Cicekli and Akman conclude that, 'The crucial point seems to be that the notion of *imitation* figures more prominently in Turing's paper than is commonly acknowledged.'[53]

Though a tiny number of commentators have come to this conclusion they have failed to recognise purpose and utility in Turing's design. Patrick Hayes and Kenneth Ford, for instance, write that 'he tells us quite clearly to make a program that can do as well as a man at pretending to be a woman … a mechanical transvestite' only to support their case that the Turing Test is a blind alley as imitating human capabilities should not be the ultimate goal of AI.[54] But once we accept the criterion with which Turing replaces the question "Can a Machine Think?" is *not* designed to ascertain whether a *machine can think* but whether a machine can produce a model of a phenomenon at least as convincing as that produced by a man, then the purpose and utility of the Test becomes immediately apparent. Indeed, as Searle observes, 'The idea that computer simulations could be the real thing ought to have seemed suspicious in the first place because the computer isn't confined to simulating mental operations by any means.'[55] Though reality can never be represented exactly by a mathematical model what we do produce is often sufficient for our purposes; and with the proliferation of cheap and increasingly sophisticated information technology the use of computers to model scientific, social and economic phenomena has in fact become increasingly prevalent. Only an obsessive preoccupation with a very specific and peculiar conception of AI can have prevented us from seeing that the Imitation Game provides us with the framework necessary for testing the success or the "intelligence" of such simulations, setting the predictions against reality first; then against the performance of proven simulators, in order to ascertain which most closely corresponds to data derived from observations in the laboratory or in the field. If there is any doubt on this score consider that Turing introduces yet another (and surprisingly overlooked) Imitation Game — in which humans do not feature at all. Proceeding on the assumption that the computer to which he has referred throughout performs satisfactorily (i.e. a Paper Machine: more on this later!) Turing suggests that his Imitation Game could even be reconfigured in a way that might permit one to test programmes on the new digital computers. 'The imitation game could then be played with the machine in question (as B) and the mimicking digital computer (as A) and the interrogator would be unable to distinguish them.'[56]

In fact, Turing's criterion for *thinking* would seem to be Poe's criterion for *analytical thought* — for the objective of the machine is to present a convincing simulation and there can be no better measure of success than deception — i.e. 'seduce into error or hurry into miscalculation'.[57] And having established this crucial point one cannot but wonder whether the huge number of conflicting opinions concerning the meaning and implications of the Test might be significant. Turing was a computer programmer, and thus entirely capable of writing instructions which are not open to misunderstanding; his paper promised a new form of the problem expressed in relatively unambiguous terms. How is the past half-century of incomprehension and interminable argument possible? The use of the misleading and loaded terms "thinking" and "thought" might be merely unfortunate. The resemblance of Turing's Test to the famous Cartesian test for *conscious thought* — in which it shown that no one could mistake a 'clockwork man' for a human with a soul — might conceivably be dismissed as a devastating blunder.[58] But how to account for his mischievous comments about the future being like the chapters describing the rise of the machines in Samuel Butler's fantasy novel *Erewhon*? '[It] seems probable that once the machine thinking method had started, it would not take long to outstrip our feeble powers', writes Turing. 'At some stage therefore we should have to expect the machines to take control'.[59] An embarrassment to proponents of Strong AI — 'this is comic-strip stuff'[60] — one begins to understand precisely why Turing was seen to giggle when he penned 'Computing Machinery and Intelligence'.[61] I believe we should at least be prepared to entertain the possibility that the enormous confusion surrounding the

Turing Test in something rather more than a mere accident. The very imprecision of the Imitation Game has helped to produce an environment in which interrogators *are* regularly seduced into serious errors regarding the ontological status of computers. Indeed, Turing's article itself might be regarded as a programme of sorts, one that has won game after game (invariably mistaken for a passage in Descartes), a mechanism for dissimulation with a track-record equal to that of the Mechanical Turk.

Once this is accepted further mysteries concerning the Imitation Game begin to fall into place. Consider the "Gödel Argument" set out by the philosopher John Lucas in his famous article 'Minds, Machines and Gödel (1961), which is developed at length by Roger Penrose in *The Emperor's New Mind* (1989) and *Shadows of the Mind* (1994). As Penrose notes, in proving no algorithm exists for deciding whether or not an algorithm run on a Turing machine is going to stop, Turing himself had shown that there can be no general algorithm for deciding mathematical questions:[62] 'By Gödel's famous theorem or some similar argument [i.e. Church-Turing!], one can show that however the machine is constructed there are bound to be cases where the machine fails to give an answer, but a mathematician would be able to'.[63] And so, Lucas and Penrose conclude, the human mind *cannot* be explained as an algorithm running on a Turing machine. Indeed, in Copeland's view, it is virtually impossible to say what mathematical conception of the mind is available to someone who endorses the Gödel Argument — 'because the objection, if sound, could be used equally well to support the conclusion, not only that the mind is not a Turing machine, but also that it is not any one of a very broad range of machines.'[64] How is this apparent contradiction in Turing's thought possible? — Penrose merely concludes that, 'There is perhaps some irony in the fact that this aspect of Turing's own work may now indirectly provide us with a possible loophole to his own viewpoint concerning the nature of mental phenomena.'[65] I would go considerable further, and suggest it is quite unthinkable that the mathematician who established the existence of non-computable numbers could have been suggesting in all seriousness that the human mind is an algorithm. In fact, Turing can be seen to have anticipated that his original and imprecise question might provoke what he termed the "Mathematical

Objection", and we are now in a position to understand his response. Citing the mathematicians who have worked for centuries on the question as to whether Fermat's last theorem is true or not, Turing jokes that, 'The short answer to this argument is that although it is established that there are limitations to the powers of any particular machine, it has only been stated, without any sort of proof, that no such limitations apply to the human intellect'.[66] But Turing insists he does not 'think this view can be dismissed quite so lightly', and his careful response is not what might be expected of a proponent of Strong AI.[67] Turing sets out procedures whereby a machine might be made to appear as though it had overcome the Decision Problem — permitted to risk mistakes for the sake of experiment and provided with a memory-bank where the results can be stored, a programme could modify itself over time, effectively becoming a different machine, thereby avoiding the problem altogether. Turing's machine would not have overcome the Mathematical Objection, but this deficiency would no longer be apparent to the average interrogator. As Turing says, his contention is merely that 'machines can be constructed which will simulate the behaviour of the human mind very closely.'[68] Evidently, the point to all this has nothing to do with substantiating the claims of Strong AI. Turing is concerned primarily with the production of a machine that will effectively *simulate* the natural phenomenon that is at once the most familiar to us and the most challenging to model, the human mind. And this may explain why Turing is prepared to countenance what he terms *faking* — predetermining some of the decisions that must be made by a machine — in order to hasten and control its education and development.[69] 'It would be quite easy to arrange the experiences in such a way that they automatically caused the structure of the machine to build up to a previously intended form, and this would obviously be a gross form of cheating, almost on a par with having a man inside the machine.'[70]

Finally, I would like to suggest we are now in a position to resolve one of the most puzzling aspects of the "Chinese Room". In this famous attack on the Turing Test and the related claims of Strong AI, Searle (1980) imagines he is locked into a room and given counters marked with Chinese symbols together with a set of instructions (in English) on how to manipulate them.[71] Searle then passes a version of the Turing Test, providing appropriate responses to questions from outside the room (written in Hanzi) merely by referring to his table of instructions — while doggedly insisting he does not understand one word of Chinese! 'For the purposes of the Chinese, I am simply an instantiation of the computer program.'[72] According to the proponents of Strong AI, Searle has proven that he understood what was written in Chinese, but to Searle it is quite obvious he has not. 'I have inputs and outputs that are indistinguishable from those of the native Chinese speaker and I can have any formal program you like, but I still understand nothing.'[73] Searle's thought-experiment has generated a vast critical literature but what is perhaps the most curious aspect would seem to have attracted relatively little comment. The fact is that, far from being a modification of the Turing Test (a subversive substitution of a philosopher for a computer) Searle has described precisely the scenario that Turing had in mind. It will be remembered that a Digital Computer is to play against another 'discrete state machine' which has already played the Imitation Game — and this could hardly be anything other than the sort of *Human Computer* Searle jokes about. 'The idea behind digital computers may be explained by saying that these machines are intended to carry out any operations which could be done by a human computer', writes Turing. 'The human computer is supposed to be following fixed rules; he has no authority to deviate from them in any detail.'[74] Elsewhere, Turing refers to this combination of a human with written instructions as a *paper machine*: 'A man provided with paper, pencil, and rubber, and subject to strict discipline, is in effect a universal machine.'[75] And it is clear that Turing's earliest thoughts concerning the Imitation Game arose from experience with the experimental chess-programmes run on such paper machines: 'One can produce "paper machines" for playing chess'.[76] Even Turing's self-adjusting algorithms, were run on *himself* and not a Digital Computer: 'I made a start on the latter but found the work altogether too laborious at present', he explains. 'When some electronic machines are in actual operation I hope that they will make this more feasible … instead of having to work with a paper machine as at present'.[77] It is safe to conclude the position on Strong AI frequently attributed to Turing can hardly have been held by a man who had worked for so many years locked up in Chinese Room. Indeed, we can begin to understand why Turing's surprise at the results produced by his machines is quite so pronounced. As the sentient CPU at the heart of such a machine, Turing would often have found himself performing operations, following his own written instructions, that he could never have anticipated. 'Certainly the machine can only do what what we do order it to perform, anything else would be a mechanical fault', he remarks. 'But there is no need to suppose that, when we give it its orders we know what we are doing, or what the consequences of these orders are going to be.' He concludes that: 'One does not need to be able to understand how these orders lead to the mechanism's subsequent behaviour, any more than one needs to understand the mechanism of germination when one puts a seed in the ground.'[78]

The analogy is suggestive. One could argue that agriculture is a sort of self-adjusting routine, run on human hardware for millennia, evolving in response to environmental pressures without much in the way of conscious innovation, producing outcomes that no one could have foreseen at inception. If we repeat the error of anthropomorphising *thought* we might speculate that while Machine Intelligence cannot be an *ego* it might nevertheless be a *person*, in the original sense of that term; a virtual entity sustained by a cultural practice — e.g. *persona ficta*. 'For as men make automata that move themselves by springs and wheels,' Hobbes declared, 'so men create that great LEVIATHAN … which is but an artificial man, in which the sovereignty is an artificial soul…'[79] This position would be congruent with what is called the Virtual Mind Reply to Searle's Chinese Room, but would only perpetuate the Promethean error we seek to disentangle. Turing himself expressed the hope that digital computing machines would eventually stimulate a considerable interest in symbolic logic and mathematical philosophy (noting that the language in which one communicates with the machines, i.e. the language of instruction tables, forms a sort of symbolic logic).[80] For this reason I would like to close this section by suggesting that the clearest conception of what such a *machine for thinking* might be in essence is to be found not in the Ultron computers of sci-fi, but in the thoroughly low-tech "Glass Bead Game" imagined by Hermann Hesse in *Das Glasperlenspiel* (1943).

> A reader who chanced to be ignorant of the Glass Bead Game might imagine such a Game pattern as rather similar to the pattern of a chess game, except that the significances of the pieces and the potentialities of their relation-ships to one another and their effect upon one another multiplied many-fold and an actual content must be ascribed to each piece, each constellation, each chess move, of which this move, configuration, and so on is the symbol.[81]

Inspired by Liebniz's unrealised fantasy of a universal language (which would eventually culminate, via that philosopher's engagement with the *I Ching* or *Book of Changes*, in the Binary Code), the Glass Bead Game is a cultural practice unique to the university-province of Castalia; a metalanguage that began as a game, but that developed over time into a sophisticated medium capable of comprehending the arts and sciences, permitting players to construct models of those thought processes through which humans interpret reality: 'a language of symbol and formulas, in which mathematics and music played an equal part, so that it became possible to combine astronomical and musical formulas, to reduce mathematics and music to a common denominator, as it were.'[82] Like Liebniz, the inventor of the game has taken inspiration from the prevalent misconception of Hanzi script as a system of ideograms; and the one glass bead game described in any detail in the book is based on the ritual Confucian pattern for the building of a house.[83] In this context, it might be amusing to relate that this "Chinese Room" is performed for the Magister Ludi by an assistant *who speaks no Chinese!* ('It was far too late for him to learn it now.'[84]) And, as with Turing's early coding, the final programme is run on a *human machine*. 'With a luminous golden stylus he delicately inscribed character after character on the small tablet before him, and the same characters promptly appeared in the script of the Game, enlarged a hundredfold, upon the gigantic board on the rear wall of the hall, to be spelled out by a thousand whispering voices, called out by the Speakers, broadcast to the country and the world.'[85]

CODA.

One might expect the *Prime Radiant*, the computer in Isaac Asimov's *Foundation Trilogy* (1942-1953) to be rather more high-tech: a near-contemporary (in real time) of the monkish Castalia, the Galactic Empire possesses nuclear power and star-cruisers and energy-shields. So it comes as something of a surprise to find (in the final part of the trilogy) that the Prime Radiant is really little more than a big white-board. 'First, a pearly white, unrelieved, then a trace of faint darkness here and there, and finally, the fine neatly printed equations in black, with an occasional red hairline that wavered through the darker forest like a staggering rillet.'[86] Since their fleeting appearance on the pocket-calculator of Hari Seldon the psychohistorian in the first book, the symbols projected by the Prime Radiant have been modified by a series of *human computers*, who have made alterations to the programme according to instructions in response to external data. 'It was a room which, through the centuries, had been the abode of pure science — yet it had none of the gadgets with which, through millennia of association, science has come to be considered equivalent', explains Asimov. 'It was a science, instead, which dealt with mathematical concepts only, in a manner similar to the speculation of ancient, ancient races in the primitive, prehistoric days before technology had come to be; before Man had spread beyond a single, now-unknown world'.[87] Like Castalia's Glass Bead Game, psychohistory is considered the culmination of a Leibnizian symbolic logic: but where the Masters of Castalia have decisively rejected that 'philosophy of history of which Hegel is the most brilliant and also most dangerous representative', psychohistorians of the Second Foundation have applied their skills in computer programming to the historical process, modelling with apparent success the entire history of a Galactic Empire over tens of thousands of years.[88] In fact, the series of crises, and the dialectical progression, predicted by psychohistorians cannot fail to recall the Historical Materialism of Marx; while the Seldon Plan to alleviate the worst of the coming catastrophe is entirely in keeping with the mania for planning that characterised the era from 1925 to 1975 (from Lenin's New Economic Plan and Le Corbusier's urbanism to the Keynesian stimulus package and China's Great Leap Forward). 'Psychohistory was the quintessence of sociology', explains Asimov; 'it was the science of human behaviour reduced to mathematical equations.'[89] This is said to be possible because, while individual human beings are unpredictable, the reaction of human mobs can be treated statistically. 'The larger the mob the greater the accuracy that could be achieved", claims Asimov. 'And the size of the human masses that Seldon worked with was no less than the population of the Galaxy, which in his time was numbered in the quintillions.'[90] If the determinism in this vision of history is (ostensibly) now a thing of the past, the suggestion that sociology might eventually come to be regarded as 'that branch of economics that deals with the reactions of human conglomerates to fixed social and economic stimuli', can be seen to have proven most prescient, anticipating the application of humanity's ever-greater processing-power to the problems of macroeconomic management over the latter half of the twentieth century.[91] 'Someday there will exist a unified social science of the kind that Asimov imagined,' insists Paul Krugman, Neo-Keynesian and Nobel Prize Winning economist in his introduction to the Folio edition of *The Foundation Trilogy* (2012), 'but for the time being economics is as close to psychohistory as you can get.[92]

This essay has shown that even Paper Machines are capable of projecting illusions that can undermine our very ontologies, and I would suggest in closing that computational technology have also worked to overwhelm our faith that human agency might shape our own history. And this is the real danger involved in computer modelling. Not that a computer-generated model might predict our future but that we permit such a model to determine our future through unquestioning faith in its veracity. Binding ourselves to a pattern of thought not reality but ratiocination, we impose restrictions on our capacity to think out new options, and to take control. And while the computer simulations favoured in the Neoliberal era, as opposed to the linear modelling that characterised the earlier Social Democratic consensus, might seem to make no claims to veracity, such simulations have proven more than capable of seducing theorists and policy-makers into serious errors too: a process documented at length by economic-historian Philip Mirowski in his book *Machine Dreams: How Economics Became a Cyborg Science* (2001). Indeed, this essay has discovered that Turing, like Poe before him, seems to have believed that *seducing into error* was nothing less than a Standard Test for any truly Analytical Engine. If simulations are likely to remain an important component in the management of our socio-political reality, Mirowski must surely be right to suggest that they will only be ‘dependably productive … when they have been developed and transformed from the status of *representations* to the status of *technologies*’.[93]

[1] Edgar Allan Poe, “Maelzel's Chess-Player”, *Southern Literary Journal*, April 1836, 2:318-326.

[2] Gerald M. Levitt, *The Turk, chess automaton* (McFarland & Co Inc Pub., 2000), 97, 27-42.

[3] Poe, 318-326.

[4] Jean Baudrillard, *The Transparency of Evil*, trans. James Benedict (Verso: New York, 1993), 68.

[5] Walter, Benjamin, ‘Theses on the Philosophy of History’ [1940], *Illuminations*, trans. Harry Zorn, (London: Random House, 1999), 245.

[6] Poe, “Maelzel's Chess-Player”, 318-326.

[7] *Ibid*., 318-326.

[8] *Ibid*., 318-326.

[9] *Ibid*., 318-326.

[10] 'The idea of a digital computer is an old one. Charles Babbage, Lucasian Professor of Methematics at Cambridge from 1828 to 1839, planned such a machine, called the Analytical Engine, but it was never completed. Although Babbage had all the essential ideas, his machine was not at that time such a very attractive prospect. The speed which would have been available would be definitely faster than a human computer but something like 100 times slower than the Manchester Machine, itself one of the slower of the modern machines. The storage was to be purely mechanical, using wheels and cards.' - Alan Turing, ’Computing Machinery and Intelligence' (1948), *The Essential Turing*, 446. Also, see his essay on 'Chess' (1953), 573.

[11] Poe, “Maelzel's Chess-Player”, 318-326.

[12] Poe, *Selected Tales* (Penguin: London, 2007), 118.

[13] *Ibid*., 350.

[14] *Ibid*., 118.

[15] *Ibid*., 119.

[16] *Ibid*.

[17] *Ibid*., 350.

[18] *Ibid*., 118.

[19] *Ibid*. 119.

[20] *Ibid*. 119

[21] *Ibid*., 119.

[22] Norbert Wiener, *Cybernetics: or Control and Communication in the Animal and the Machine* (MIT press, 1948, 1961), 173.

[23] Turing, 570. Wiener, 164. Poe, 348.

[24] Wiener, 165.

[25] *Ibid*., 172.

[26] *Ibid*., 172.

[27] B. Jack Copeland, *The Essential Turing*, 566.

[28] Noam Chomsky, *Language and Thought* (London: Moyer Bell, 1993), 93.

[29] Copeland, 566.

[30] Robert Wright, 'Can Machines Think?', *Time Magazine* (25 March 1996) — available online at: http://content.time.com/time/subscriber/article/0,33009,984304-1,00.html

[31] Amrose Bierce, 'Moxon's Monster', [1893], *Can Such Things Be?* (1st World Books: Fairfield, IA, 2006), 65.

[32] *Ibid*.

[33] Sigmund Freud, 'The Uncanny' [1919], *The Uncanny*, trans. David McLintock (Penguin: London, 2003), 135.

[34] *Ibid*., 141.

[35] *Ibid*., 142.

[36] Turing, 431.

[37] See Turing's essay on 'Chess' (1953), and also B. Jack Copeland's introduction, in *The Essential Turing* (Oxford: Clarendon Press):

[38] Turing, 'Intelligent Machinery' (1948), 412.

[39] Alan Turing, 'Computing Machinery and Intelligence' (1950). First published in *Mind*, 59 (1950); 433-60. Republished, B. Jack Copeland, ed., *The Essential Turing* (Oxford: Clarendon Press, 2004), 441.

[40] *Ibid*.

[41] *Ibid*.

[42] *Ibid*.

[43] *Ibid*.

[44] *Ibid*.

[45] See S.G. Sterrett, 'Turing's Two Tests for Intelligence', *Minds and Machines* 10 (2000), 541-59. Reprinted J.H. Moor (ed.), *The Turing Test* (Dordrecht: Kluwer, 2003).

[46] Turing, 495, 485.

[47] Roger Penrose, *The Emperor's New Mind* (Oxford: OUP, 1989, 1999), 7.

[48] Copeland, 435.

[49] Andrew Hodges, *Alan Turing: the Enigma* (1983), 415. Copeland, 435.

[50] Ayse Pinar Saygin, Ilyas Cicekli, Varol Akman, 'Turing Test: 50 Years Later', *Minds and Machines* 10 (2000); 467.

[51] *Ibid*.

[52] *Ibid*.

[53] *Ibid*.

[54] See Judith Genova, 'Turing's Sexual Guessing Game', *Social Epistemology* 8:4 (1994); 313-326. Also, Patrick Hayes and Kenneth Ford, 'Turing Test Considered Harmful', *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence* (1995); 972-977.

[55] John R. Searle, 'Minds, Brains and Programs', *Behavioural and Brain Sciences* 3:3 (1980): 417-457.

[56] Turing, 448.

[57] Poe, 'Murders in the Rue Morgue', 119.

[58] Rene Descartes, *Discourse on Method and Other Writings*, trans. F.E. Sutcliffe (London: Penguin, 1968), 103.

[59] Turing, 475.

[60] Copeland, 470.

[61] 'I remember him reading aloud to me some of the passages — always with a smile, sometimes with a giggle.' — R. Gandy, 'Human versus Mechancial Intelligence', in P. Millican and A. Clark (eds.) *Machines and Thought: The Legacy of Alan Turing*, vol. i (Oxford: Clarendon Press, 1996), 125.

[62] Penrose, 83.

[63] Turing, 473.

[64] Copeland, 468.

[65] Roger Penrose, *The Emperor's New Mind* (Oxford University Press, 1989, 1999), 46.

[66] Turing, 451.

[67] *Ibid*.

[68] *Ibid*., 473.

[69] *Ibid*., 475.

[70] *Ibid*., 473.

[71] Searle, 417-457.

[72] *Ibid*., 420.

[73] *Ibid*.

[74] Turing, 444.

[75] *Ibid*., 416.

[76] *Ibid*., 412.

[77] *Ibid*., 428.

[78] Ibid., 485.

[79] Thomas Hobbes, *Leviathan* (1651)

[80] Turing, 392.

[81] Hesse, *The Glass Bead Game*, 114.

[82] *Ibid*., 28.

[83] *Ibid*., 230.

[84] *Ibid*., 230.

[85] *Ibid*, 248.

[86] Isaac Asimov, *Foundation Trilogy*, 494.

[87] Asimov, 493.

[88] Hesse, 331.

[89] Asimov, 411.

[90] Asimov, 411.

[91] Asimov, 17.

[92] Paul Krugman, Incidents from my career. http://web.mit.edu/krugman/www/incidents.html

[93] *Ibid*., 532.