THE FUNDAMENTAL DISTINCTION BETWEEN BRAINS AND TURING MACHINES

By Andrew Friedman

The exponential growth in computing power in the past few decades has led to the genuine reality that computers have surpassed human intelligence in many realms. This fact, along with other observations such as the apparent similarity between binary arithmetic and the behavior of neurons, and the digital nature of DNA, has led to the fairly widespread belief that the brain is just an organic computer, albeit a highly complex one. The logical extension of the idea of the brain as a computer is that one day, given sufficient complexity, computers will not only surpass us in computational ability, but will also achieve the status of conscious beings, just like us. This belief of nearly inevitable conscious machines has become ingrained in the public consciousness through popular depictions of androids and other conscious robots in literature and film. And indeed it is a belief that is taken seriously by many in the artificial intelligence community. As such, the purpose of this article is to examine, in detail, whether such a belief is justified.

To examine whether the brain is just an organic computer, it becomes necessary to start from the theoretical definition of a computer, the Turing Machine, first envisioned by Alan Turing, whose work in the early 1950's formed the foundation for the theory of modern day computer science. Mathematically, a Turing Machine is an abstract algorithmic manipulator of intrinsically meaningless symbols, usually the binary digits, or bits, 0 and 1. The idealized Turing Machine consists of an infinite strip of tape made of squares containing 1's or 0's, a reading head that reads the bit in the present square, a track where the reading head can move between squares in both directions, and a writing instrument that can change the bit in a square from 1 to 0 or vice versa. Algorithms, or programs, are enacted by some sequence of these basic operations, of which there turn out to be only 7: read 1, read 0, write 1, write 0, move left to position i, move right to position j, and stop. (1) Despite the high level of abstraction involved in this concept, all conventional computer software and indeed all standard home computers, from Macs to PCs, can be described in this framework as algorithms implemented on a Turing Machine. In other words, no matter how complex you get with, say, a Pentium IV PC, 1200 GHz processor running Photoshop 6.0 and 9 other programs simultaneously, when you look at it from a fundamental level, it is just a Turing Machine running a bunch of algorithms.

Given this notion of a Turing Machine, the central question of this article then becomes, "*Are our brains merely Turing Machines, where our conscious minds are simply algorithmic programs?*" This article will argue that, in fact, our brains are fundamentally different than Turing Machines, and that a Turing Machine itself could never be conscious in principle. This does not rule out the possibility of consciousness involving hardware other than neurons, or even the possibility of consciousness that we helped to create, only that consciousness itself can never be implemented on a Turing Machine alone. To begin, it will be useful to discuss how one might test in practice whether existing or possible future Turing Machines are conscious. From this, we can extend the notion of practical tests for consciousness to stronger theoretical arguments for why conscious minds can never be simply Turing Machines. But first, we must go to the Turing Test.

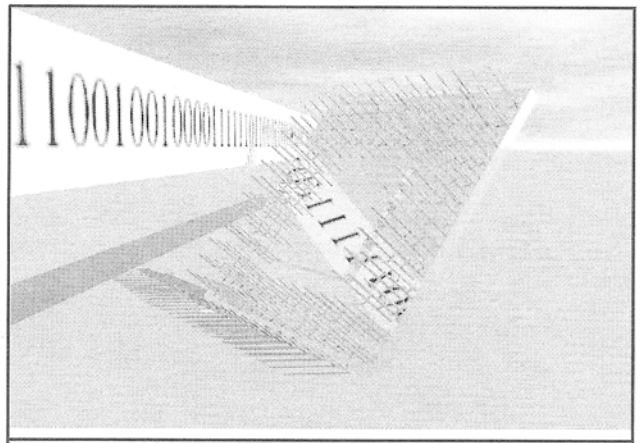

Figure 1: The idealized Turing Machine tape, stretching off to infinity, reflected here in an object that is the physical embodiment of real Turing Machines, the microprocessor.

"Seated where he could catch the readings on the two gauges of the Voight-Kampff testing apparatus, Rick Deckard said, I'm going to outline a number of social situations. You are to express your reaction to each as quickly as possible. You will be timed, of course."

"And of course," Rachel said distantly, "my verbal responses won't count. It's solely the eye-muscle and capillary reaction that you'll use as indices."

--- Philip K. Dick, "Do Androids Dream of Electric Sheep?" (2)

Implicit here in this Philip K. Dick passage is a version of the Turing Test for artificial intelligence. Turing himself was thinking about the interesting question of how one might be able to tell the difference between a conscious human being and a sufficiently advanced Turing Machine programmed to mimic human behavior down to the finest detail. The place to start, Turing thought, would be to examine the behavior of both the machine and a human subject in a setting where the observer does not know in advance which is which. The subjects are then given a test of some sort designed to identify attributes that are uniquely human, often via an anonymous text based interface. Maybe the test looks at linguistic ability, or creative processes, or the ability to perform higher reasoning. In any case, after the test, if the observer says subject A is human, then subject A has passed the Turing Test, and we are to believe that subject A is a conscious, intelligent human being.

The idea of the Turing Test can be generalized to any situation, however broad or limited we like, where an uninformed observer simply has to identify the true conscious beings from amongst the fakers. To check the observer's accuracy, one could construct tests where all subjects are machines or all subjects are humans, with most tests having some statistically significant mixture thereof. One could even instruct the humans to behave deliberately like a machine, to try to throw the observer for a loop. The test also need not be limited to a text-based interface, where appearances are neglected. At the far extreme, it could include the testing of an android supposedly capable of completely mimicking human behavior, and passing off as a full-fledged member of human society. This is exactly the situation inherent in the aforementioned Philip K. Dick passage.

As it was, Turing himself did not envision this idea as a rigorous test, only as a first approximation, which would be the obvious thing to do. Nevertheless, many have adopted the Turing Test as a seemingly objective test for consciousness, and have used examples such as conversation programs and chess playing programs, that have already passed limited Turing Tests, as evidence for the inevitability of future conscious

machines. I will argue here that this is an unreasonable position, and that the utility of the Turing Test is severely limited regarding the fundamental question of consciousness.

As perhaps the most striking example, in 1997, Deep Blue, the chess playing computer, defeated chess grandmaster Gary Kasparov. (3) In this sense, in the context of chess playing, the Turing Machine Deep Blue has passed a limited Turing Test. Yet we know with certainty that Deep Blue is not conscious. The team of engineers and computer scientists that built Deep Blue can reconstruct the operations enacted at every step in the program, and there is no mystery to the fact that Deep Blue is just an unconscious algorithm running on standard silicon hardware. Thus, at the extreme, an observer taking the Turing Test at face value would be fooled into thinking that, while it is playing chess, Deep Blue is actually a conscious thinking being.

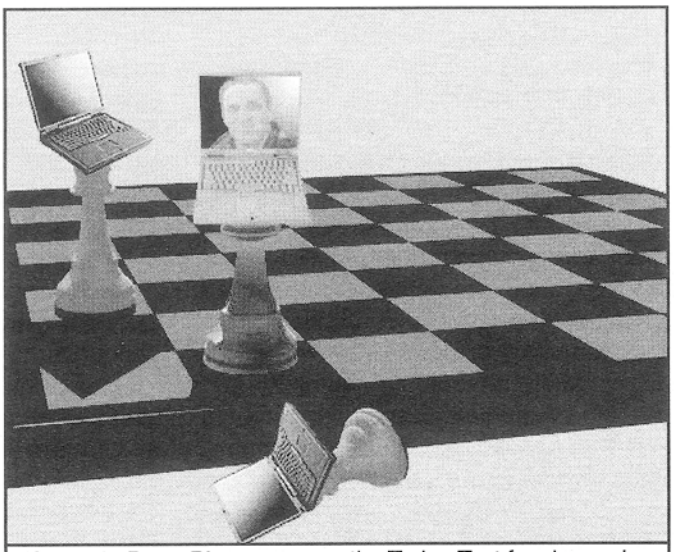

Figure 2: Deep Blue may pass the Turing Test for chess playing, but only Kasparov is a true conscious being. To put them on the same footing regarding consciousness would be as ridiculous as this absurd representation of their chess battle.

This is not to say that anyone is seriously equating chess playing with consciousness, only that some view the event as indicative that given sufficient complexity, some future variation of Deep Blue will indeed be conscious. As it is, the path that may lead one to extrapolate from Deep Blue to Commander Data from *Star Trek* with an emotion chip comes largely from the irrational, visceral response one experiences when we hear that a lump of metal and plastic beat our man Kasparov. First of all, we should in no way be surprised that machines can surpass human intelligence, which itself is distinct from consciousness. Pocket calculators have long since exceeded human computational ability, and seriously, computer games having been beating players at chess for years. Here it is only that this was the first situation where the best human did not beat the best computer in chess, and as such it has been construed by many as a surprising blow to the pride of the species.

The point is that while passing a Turing Test may be necessary for consciousness, it is clearly not sufficient, as Deep Blue's success still does not imply that it is conscious. Furthermore, the passage of a Turing Test may not even be necessary for consciousness, as it is quite conceivable that under certain conditions, even a conscious human observer could fail! Consider an autistic child or a purposely machine-like actor as Turing Test subjects, for example. Certainly one would require Turing Tests of broader scope than chess playing for an observer to make a more meaningful assessment of consciousness, but in nearly all such tests imaginable, it is hard not to conceive of a way in which the observer could be fooled.

So what is the bottom line with a Turing Test for consciousness? Often, it tells you something meaningful when a subject fails a Turing Test, namely that the subject is truly not a conscious being. Although this is not always accurate, because, as we just discussed, even humans can sometimes fail the Turing Test. And even when a subject passes a Turing Test, it still does not tell you very much. When a human passes, you have simply confirmed the obvious, and when a machine passes, as has already happened in limited cases, you have simply demonstrated the extent to which the observer can be fooled. At heart, the Turing Test is more useful as a measure of how stupid we are as observers. At the extreme, a scarecrow, or a mannequin at 100 yards could pass a Turing Test, just as the audio-animatronic exhibits at Disneyland can easily convince most five-year-olds.

"At the extreme, a scarecrow, or a mannequin at 100 yards could pass a Turing Test...."

Furthermore, all Turing Tests are limited. Even if a computer can run a program that simulates pleasant, thoughtful conversation, demonstrating a respectable amount of wit, the fact that it can fool us simply does not imply that the program is actually conscious. It is perfectly conceivable that we could build an unconscious machine so completely convincing at approximating conscious behavior that no human observer could ever tell the difference. But until we can actually build a machine that behaves like Commander Data, the importance of machines passing Turing Tests will be separated only by a matter of degree from those with mannequins and scarecrows.

Excessive importance attached to the Turing Test seems to be merely a relic of the now defunct psychological school of behaviorism, where internal states are completely ignored and conscious entities are treated simply in terms of their input and output states. Since we now have the tools in computer science and neurobiology to talk about the internal states of both computers and the brain, it would be bordering on ridiculous to regress to the behaviorist paradigm. The purpose of this critique of the Turing Test is not to imply that it is used as dogma in the artificial intelligence community, which it certainly is not. The major goal here was simply to illustrate how easily one can be misled by the importance of a machine passing a Turing Test. Regarding consciousness, one should view such a success with due skepticism.

In the end, we are asking whether the Turing Test tells us anything about whether a machine is actually conscious. From the above arguments, we conclude that in this regard, the Turing Test tells us almost nothing. This leads us to the notion that practical tests for consciousness such as the Turing Test, although interesting, may not be of much use towards answering the fundamental question of whether or not brains are just Turing Machines. Thus it becomes more useful to examine the question from a more theoretical standpoint.

If a particular Turing Machine were truly conscious, one might reasonably ask the fundamental theoretical question, "*What are the necessary and sufficient conditions for consciousness, and how, specifically, are they being met by this hypothetical Turing Machine?*" As it stands, the philosophical position of functionalism, or what the philosopher John Searle calls Strong Artificial Intelligence (Strong AI), is quite precise about the necessary and sufficient conditions for consciousness. Its basic tenet is that you need physical hardware capable of implementing a sufficiently complex algorithm son a Turing Machine, and once the proper algorithm is running, you will have consciousness. The Mathematical Physicist Roger Penrose states the Strong AI position thus.

"According to Strong AI, it is simply the algorithm that counts. It makes no difference whether that algorithm is being effected by a brain, an electronic computer, an entire country of Indians, a mechanical device of wheels and cogs, or a system of water pipes. The viewpoint is that it is simply the logical structure of the algorithm that is significant for the mental state it is supposed to represent." (4)

Regarding how complex this algorithm need be, as the Theoretical Astrophysicist Max Tegmark notes, the requirement that the mind, "possess a certain minimum complexity goes without saying. In this vein, Barrow has suggested that only structures complex enough for Godel's Incompleteness Theorem to apply can contain [conscious beings.]" (5)

Godel's Theorem, one of the major discoveries of 20th century mathematics, essentially states that any formal mathematical system of symbols and rules that is complex enough to encode the operations of arithmetic (i.e. addition, subtraction, multiplication, etc...), can either be **consistent** or **complete**, but never both simultaneously. In this context, a formal system is **consistent** if it leads to no contradictions, where say, some statement can be proven to be both true and false simultaneously. On the other hand, a system is **complete** if all true and false statements could be proven to be true or false within that system. At first glance, one would think that any worthwhile system would have both of these properties, but as it turns out, nature is not so kind when it comes to systems that are complex enough to encode something so seemingly basic as arithmetic. The upshot is that for every consistent formal system complex enough to encode arithmetic, there will be true and false statements that will not be provable within that system. So what does this foray into the nature of mathematical proof have to do with our theoretical discussion of consciousness and Turing Machines? As it turns out, quite a bit, as we will discuss later. But for now, it will be good enough to think of Godel's Theorem and what it implies as a kind of hypothetical benchmark for the minimum level of complexity a Turing Machine would need to have to be conscious. Anything less complex, such as a rock, for example, simply would not have what it takes.

Thus, the necessary condition of Godel complexity (6) seems as reasonable a cut off point as we can get for consciousness, but as Tegmark agrees, "this is of course, unlikely to be a sufficient condition." (7) To be fair, there is no exact Strong AI consensus as to the required level of complexity. A level higher than mere Godel complexity may even be required to guarantee consciousness. But the Strong AI position does claim that, independent of the hardware, some level of algorithmic complexity implemented on a Turing Machine will be both necessary and sufficient for consciousness. In contrast, this thesis holds that algorithmic programming alone is not sufficient for consciousness. The reason for this can be succinctly illustrated by a three-step argument set forth by John Searle, which hinges on the differences between syntax, the bare logical structure of a sentence comprised of symbols, and semantics, the actual meaning contained in that sentence. Searle argues that:

1.Programs are entirely syntactical.
2.Minds have semantic contents.
3.Syntax is not the same as, nor by itself sufficient for semantics. Therefore programs are not minds. (8)

The power of this argument is illustrated by yet another argument of Searle's, his famous Chinese room argument. (Refer also to the Searle Interview.)

The Chinese Room Argument asks us to consider a scenario where Searle is trapped in a room where messages written in Chinese are passed to him through a slot. The messages ask certain questions which Searle must provide an answer to in the form of another message written in Chinese which he passes back through the slot. To do this, he has a rule book which tells him how to manipulate the Chinese symbols in the message in order to construct the appropriate answer, also comprised of Chinese symbols. What Searle argues is that given a good enough rule book, or algorithm, he could provide correct answers to these questions without understanding a word of Chinese! In other words, Searle's answers will always have the right syntax, but no semantics whatsoever.

"...if there is a single property, even a trivially unimportant one, that we have but Turing Machines do not have, then we can not be simply Turing Machines."

The reader may recognize the Chinese Room scenario as yet another example of a Turing Test for consciousness. Here Searle himself is performing the actions of a Turing Machine, an algorithmic manipulator of intrinsically meaningless symbols. In this scenario, we can see that, regarding the limited Turing Test involving the comprehension and composition of Chinese language, in principle, Searle can pass the test without any conscious understanding of Chinese. What Searle argues is that if he does not understand Chinese while enacting the algorithm, why on earth should we believe that a Turing Machine understands it? Searle then ask us, since according to Strong AI, the algorithm is all that counts, and here he is performing the algorithm, what does a Turing Machine have that he does not have in this scenario? The answer is that it has nothing that he does not have. In fact, they are both enacting the same algorithm. In the end, through arguments such as this, we see there are strong theoretical reasons for believing that complex algorithmic programs alone can never be sufficient for consciousness. (9)

So if algorithmic complexity alone is not sufficient for consciousness, what is? In other words, given the fact that we are conscious beings, what do we have that Turing Machines do not have? This question hits directly upon the central point of this article, which says that if there is a single property, even a trivially unimportant one, that we have but Turing Machines do not have, then we can not be simply Turing Machines. In the rest of this article I shall discuss many such properties, the most important of which is the fact that human brains evolve through random biological evolution, while Turing Machines need not come into existence that way. In fact, the thesis here claims that random evolutionary mechanisms are always necessary requirements for consciousness to arise, and that in total, Godel complexity and sufficient time for random evolution form the necessary and sufficient conditions for consciousness. (10)

Regarding evolution, it is of interest to note that in Searle's Chinese Room scenario, Searle could genuinely pass the Turing Test if he was allowed to evolve by natural selection through random trial and error guesswork in such a way as to actually learn Chinese. Such an outcome would undoubtedly require a more rigorous and self-referential program to work with, and would be nearly impossible in practice, but not in principle. Some have noted the practical difficulty of Searle actually carrying out the algorithm and have envisioned scenarios that replace Searle with the entire population of India, as to make the program's implementation more tractable. An interesting question then becomes, what exactly would it mean for the entire country to genuinely pass a Turing Test for consciousness? It would mean nothing other than the evolution of a telepathic collective consciousness amongst the country's members. Such suppositions might be dismissed offhand, if they were not so close to the accepted Strong AI position, which holds that the country of India would be "conscious"merely by implementing the proper algorithm. Standard Strong AI simply does not use the word telepathy, while its claims are actually more unlikely, considering that such a version of a "conscious India" could occur without any kind of evolution.

That aside, since we know that human minds are a result of biological evolution, it seems that the Strong AI adherent would have to hold that if our minds are Turing Machines, then they were simply programmed through evolution. Unfortunately, such a view is severely in error, primarily because random evolution is almost the pure antithesis of direct programming. The key difference here is that evolution relies on random mechanisms such as genetic mutations, whereas the act of programming itself is not random. Clearly, programmers do not design software through a series of coin flips. (11) Thus, we differ from Turing Machines at least in regard to how we were created.

One might argue erroneously that the programmer could actually implement the effects of random evolution simply by programming randomness. Bennett addresses this when she writes of attempts to generate random sequences on computers,

"Within any sequence generated by the computer through a programmed algorithm or formula, the next digit is a completely deterministic choice, not random in the sense that a dice throw ...or even the infinite digits of the mysterious pi are random." As a result, the digits are, *"...called pseudo-random numbers...for how could a truly random number be generated through a formula and a machine?" (12)*

When a computer tries to generate a random sequence, it may appear successful at first, but if one waits long enough, eventually the sequence will repeat, destroying the facade of randomness. (13) Genuinely random numbers are simply not computable. In this way it can be said that a Turing Machine can approximate randomness, but it can never actually compute the real thing. Therefore, true randomness could never be programmed in an attempt to mimic actual evolution. Since we know with certainty of only one example of conscious minds, which evolved through random evolutionary mechanisms, namely us, the onus then falls upon the adherent of Strong AI to show that conscious minds can arise from anything other than random processes.

Granting that random processes are required, one might argue that a programmer could simply set up the Turing Machine with the ability to reprogram itself, put it into an environment where it must learn and adapt, and wait for some sufficiently long time, thus allowing it to evolve like humans did. Such has even been attempted to some degree in modern artificial intelligence research through neural networks and genetic algorithms. (14) If a programmer wants true evolution, they must first supply the Turing Machine with a minimum level of complexity, place it in some appropriate environment, and then just let it go, avoiding interference at all costs. Otherwise, they would compromise the random nature of the evolutionary process with every act of programmed interference. In fact, this is precisely what must be done if humans ever hope to aid in the creation of non-human consciousness. And in fact, it is likely that at some point in the past, the genetic ancestors of humans were mere Turing Machines! But the key point here is that, after the Turing Machine has evolved beyond a certain point, it will necessarily no longer be a Turing Machine.

One way of quantifying this statement mathematically is to note that Turing Machines have at most a countably infinite set of operations, which tells us that they can only approximate, but never truly compute a random operation. An "Evolved Turing Machine" which could "compute" randomness, therefore would have to have at least a continuously infinite set of operations. For those who are familiar with set theory, this is simply the statement that Turing Machines can have at most the cardinality, or "size of infinity", of the integers, whereas something capable of true randomness would need to have at least the cardinality of the Real Numbers, which have a larger infinity than the integers. (15) It is intriguing that the seemingly obscure mathematical fact that there are different sizes of infinity would have bearing on the theoretical notion of consciousness, but it does, and it can not be ignored. If human beings have even the tiniest ability to perform actions and generate thoughts which are truly random, then the cardinality of our minds necessarily exceeds that of any Turing Machine. Therefore, we cannot be defined as mere Turing Machines.

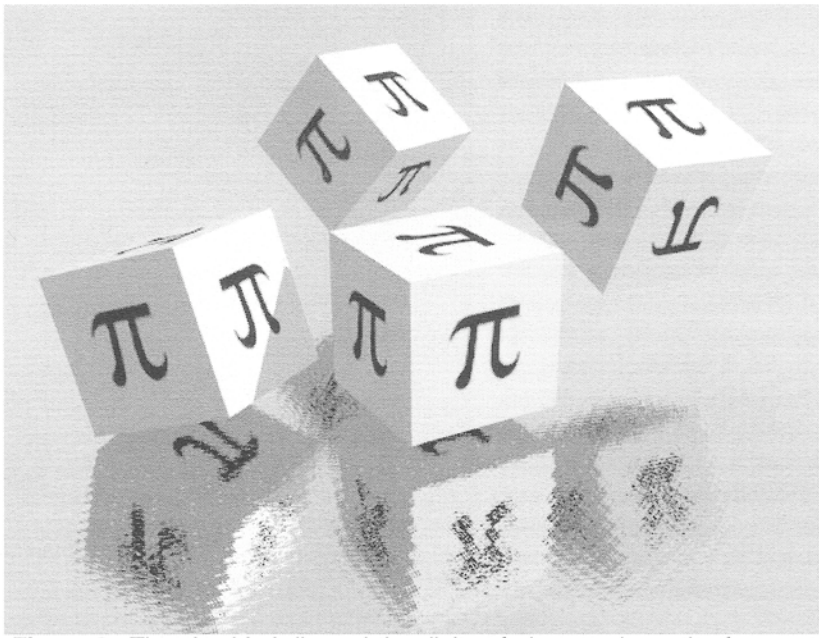

Figure 3: The six sided die and the digits of pi, as perhaps the foremost symbols of randomness, seemed quite appropriate to combine here. It may be that our basic creative processes, such as the very ones used in creating this graphic, require uncomputable, random mechanisms that are unique to human beings, and thus could only be simulated on a computer.

As discussed earlier, the original Turing Machine can only perform 7 basic operations. The claim here is that, after random evolution has proceeded up to a certain point, as it has for human minds, a new randomizing operation has been effectively added to the Turing Machine. As we have seen from the uncomputability of random numbers, the randomizing operation can not be programmed into an algorithm using only the 7 basic operations. So unlike Turing Machines, because humans evolved from random processes, evidently, we do retain some ability to use genuine randomness as the carry-over of our genetic history. (16) Although we usually only see randomness in the context of human creativity, inconsistency, and irrational behavior, in principle, it could conceivably be used to construct a genuinely random sequence of arbitrary length. (17) The claim here is simply that a human could act as a fair coin while, even in principle, a Turing Machine could at best approximate one. Again, the relevant point is that if humans have any truly random ability whatsoever, then we are different from Turing Machines.

But it is important to note that human minds are certainly not entirely random. First of all, few would argue with the claim that nearly all of our unconscious processes (18) are completely algorithmic and could be implemented on a Turing Machine. Kidney regulation, the operation of our pancreas, and immune response are undoubtedly governed by complex brain processes, but those processes are almost certainly only programs. This said, even the fiercest opponents of Strong AI must grant that at least part of our brain is simply an organic computer. Unfortunately, this suggestive similarity has led many adherents of Strong AI to suggest that the entire brain

Figure 4: Mathematicians represent the different sizes of infinity with the Hebrew letter Aleph and the subscripts 1 and 0, where Aleph 1 represents the "larger" infinity of the real numbers, and Aleph 0 represents the standard, "smaller" infinity of the integers. What is relevant here is that while Turing Machines are most certainly limited to the smaller infinity, the minds of human beings may not be, making us fundamentally different from computers.

is nothing more than a complex biological Turing Machine. This sort of logical leap is unjustified. The key point is that while brains and Turing Machines may share many similarities, and while *part* of the brain probably is a Turing Machine, if we can demonstrate even one difference between them, then brains cannot be entirely Turing Machines. Thus even though unconscious processes do appear to be algorithmic, the part of the mind we are actually interested in is the set of all conscious processes. It is then pertinent to show that, in regard to the most interesting mental phenomena such as irrational behavior, creativity, and language, the relevant part of the brain cannot be merely implementing algorithms. Unconscious processes aside, it is with the conscious processes that we differ fundamentally from Turing Machines.

One such conscious process that has been cited, which as promised, draws upon our earlier discussion of Godel's Theorem, is human insight into mathematical truth. In 1961, John Lucas constructed such an argument using the implications of Godel's Incompleteness Theorem in an effort to prove that minds cannot be Turing Machines; this argument has recently been revived by Roger Penrose. (19) The argument has its flaws, but it begins with a strong, unassailable point which we have already discussed. If we want to show that minds are different from Turing Machines, all we must do is demonstrate any task, even a trivially unimportant one, that we can do but a Turing Machine cannot. Casti recounts Lucas' argument.

"By standing outside the incomplete, consistent formal system Godel's results imply that humans can know there exists some true, but unprovable, statement. But the machine cannot prove this fact; hence, a human can beat every machine since such a true, but unprovable, statement exists for every machine." (20)

From this, Lucas and Penrose would have us conclude that, at least during the understanding the truth of Godel statements, human minds could not be simply implementing algorithms.

Paul Benacerraf has criticized this argument by stating that, "...Lucas has too limited a view of machines, since any machine that could reprogram itself would be exempt from the Godel argument." (21) This point is crucial as it is precisely in accordance with my present thesis, where evolution by natural selection is necessary for consciousness. In this case, evolution is made possible by allowing the machine to adapt and reprogram itself. But evolution aside, the fundamental problem with Lucas's argument is that, although the Turing Machine cannot prove a Godel statement G from its axioms, *neither could any human*! In order to known that G is true, a human must go outside the system, transcending the method of proof from within the system. The Godel proof as such, must be conducted in the meta-language of the formal system in question.

What Lucas has shown is that it is always possible for a human to speak in the meta-language of any Turing Machine. But in trying to demonstrate a task we can perform that a Turing Machine cannot, he confuses the fact that the tasks in his argument are actually different tasks. A human must "know" the truth of a G statement in the Turing Machine's formal system, while the Turing Machine is being asked to "axiomatically prove" the G statement in its own formal system.

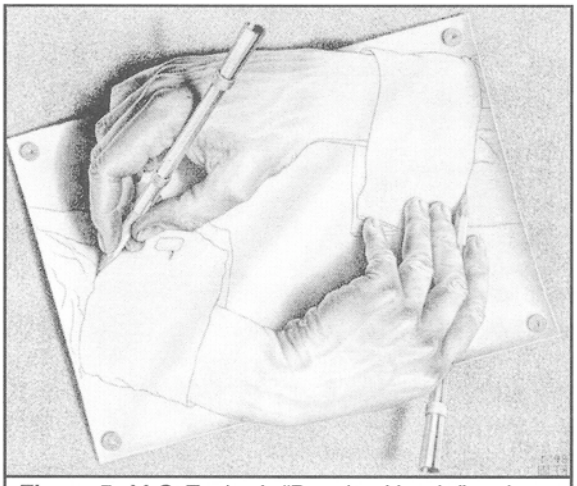

Figure 5: M.C. Escher's "Drawing Hands," perhaps the most recognizable example of visual paradox. (c) 2002 Cordon Art - Baarn - Holland. All rights reserved. Used by permission.

What then, might be substituted into the argument as genuinely equivalent tasks? One could ask, while humans can speak in their own meta-language, as demonstrated by Godel's meta-proof, can a Turing Machine generate programs that allow it to speak in its own meta-language? Well, we know for sure that it cannot possibly have that information in its repertoire to begin with. So it could speak in its own meta-language only if it were allowed to adapt and reprogram itself. But by our earlier arguments, by doing this, it would no longer be a Turing Machine. In this sense, there are no Turing Machines, no matter how complex, that could ever be able to go beyond themselves in this manner without reprogramming themselves or being reprogrammed by us. This granted, it becomes pertinent to demonstrate some examples of humans speaking in their own meta-language.

Such human meta-language fluency is evidenced by our ability to construct the Godel proof, and by similar examples of resolving paradoxes by going outside the system. One notable example of human visual paradox is represented by M.C. Escher's print, "Drawing Hands", where the question, "which hand draws which?" seems confounding at first. But as the computer scientist Douglas Hofstadter points out, the paradox is easily resolved when one ascends to the meta-level where one can clearly claim, "Escher drew both hands." (22) Similarly, when one confronts the liar paradox, in the form of the sentence, L, which says, "This sentence is false," there are several ways to resolve the issue from the meta-level where we examine the nature of the speaker of the sentence. Maybe the speaker does not know English, maybe they are mentally ill, maybe the sentence was taken out of context, maybe it was constructed randomly, or maybe the speaker is simply joking. All such observations resolve the paradox by demonstrating the ill-posed nature of the question, "Is L true or false," within the relevant system. (23) In some way, this method seems to skirt the issue, but these meta-methods are not really cheating. They simply tell us that, even though we often feel like we can understand the meaning of paradoxical statements in a given system, certain questions about them are revealed to be ill posed in that system when viewed from the meta-perspective. The claim is that while humans can always speak in their own meta-language on their own, Turing Machines can only do it if they are programmed to do so by someone else, further differentiating humans from Turing Machines.

In fact, the ability to even generate such paradoxical, and possibly inconsistent sentences, hints at another key difference between minds and Turing Machines, namely that human minds may, in fact, be inconsistent. (24) This inconsistency can be shown rather poignantly via another critique of the Lucas argument. As it is, since Lucas assumes the mind to be consistent, C.H. Whitley constructed a paradox that challenges this assumption. Whitley asked us to consider the sentence, "Lucas cannot consistently assert this sentence." This sentence is used to point out that "if Lucas could assert it, then that fact would undermine his assumed consistency." (25) Even though inconsistency would seem to undermine Lucas's argument, he might actually have had more success by basing his argument around it, since the conclusion that minds are not Turing Machines remains the same.

The upshot is that if the human mind is, in fact, inconsistent, then Godel's Incompleteness Theorem *does not apply to us*, since it only applies to consistent formal systems. (26) Regarding the assumption that minds are consistent, Barrow puts it quite succinctly.

There is really no reason to believe this and many reasons not to! The brain is a staging point in an ongoing evolutionary process. Like most evolutionary products, it does not need to be perfect, merely sufficiently good to allow selective advantage. If we admit that the mind is fallible, then the assessment of Godel sentences is beside the point. We would need to conclude that the mind was inconsistent rather than incomplete. As a result there is nothing more to be said with regard to parity with algorithmic machines. (27)

In the end, if human minds are inconsistent, at the small price of occasional calculation mistakes, inconsistencies, and flashes of irrationality, it appears that human minds may thus free themselves from the intellectual shackles of Godel's Theorem.

In conclusion, what can we say about the possibility of minds being Turing Machines? First of all, we have argued that a Turing Test can never serve as an objective measure of consciousness, as the admitted fallibility of humans simply makes us poor judges of such a subjective test. The upshot is that, like the creations of skilled magicians, Turing Machines can surely be programmed to approximate human behavior well enough to fool us, but this in no way guarantees they are conscious minds. We have also argued from a theoretical standpoint that there are several fundamental differences between minds and Turing Machines. First of all, human minds evolve through random evolutionary mechanisms, while Turing Machines do not require evolution. In addition, human minds probably possess some degree of true randomness, a quantity that is inherently uncomputable. We have also argued that while humans can speak in their own meta-language unaided, Turing Machines can never do this without reprogramming themselves or being reprogrammed. Furthermore, it is probably true that human minds, especially in regard to language, are actually inconsistent formal systems, while Turing Machines are by definition isomorphic to consistent formal systems. From all this, the logical conclusion is that minds cannot be only Turing Machines.

Strong AI is indeed correct about minds having no "preference" for biological hardware, but it is mistaken in taking the hardware/software analogy too far. In order to be consistent with the only example of consciousness we know of for certain, random evolutionary mechanisms must be taken into account. This thesis still retains the materialist part of Strong AI, which holds that minds are hardware independent, but it rejects the computational part, which holds that minds are isomorphic to Turing Machines. This is an important point, and due to the many admittedly suggestive similarities between brains and computers, it is often misunderstood. While the brain may be partly a Turing Machine, this in no way guarantees that the entire brain is just a Turing Machine. In the end, this thesis claims that conscious minds cannot arise without random evolutionary mechanisms and thus could never be products of deterministic, algorithmic programming. To put it simply, through evolution, Philip K. Dick's conscious androids with semiconductor circuit brains are indeed possible, and they would certainly be machines, but just like you and me, they would necessarily not be Turing Machines. BSJ

Notes and References

(1) Casti, John L. Casti, John L. "Five Golden Rules: Great Theories of 20th Century Mathematics: and Why They Matter" John Wiley & Sons, Inc. New York, 1996., pg. 141.

(2) Dick, Philip K. "Do Androids Dream of Electric Sheep?" , Millenium, Great Britain, 1968. pg. 44.

(3) Shaeffer, Jonathan and Plaat, Aske, "*Kasparov versus Deep Blue: The Re-match*", ICCA Journal, vol. 20, no. 2, 1997, pp. 95 - 102."
http://www.dcs.qmw.ac.uk/~icca/journal.htm

(4) Penrose, Roger. "The Emperor's New Mind." Oxford University Press, 1989. pg. 26-27.

(5) Tegmark, Max. "*Is the itheory of everythingî merely the ultimate ensemble theory?*", Annals of Physics, 270, 1998, pp. 13. Godel's Theorem states that, "For every consistent formalization of arithmetic, there exist arithmetic truths that are not provable within that formal system." Casti, pg. 163.

(6) An algorithm with Godel complexity must be isomorphic to a formal system complex enough to contain arithmetic. In other words, it must be capable of making self-referential statements. See Casti, Ch. 4.

(7) Tegmark, pg. 13.

(8) Searle, John. "The Mystery of Consciousness", The New York Review of Books, 1997. pg. 191, 11.

(9) Searle, Ch 1-4.

(10) According to this criterion for consciousness, things that do not have Godel complexity, exist fleetingly, or do not evolve, can not be conscious. In this manner we avoid panpsychism, as rocks, clouds, and individual gas molecules are not accorded consciousness, although lower animals are certain to have some measure of consciousness, and could presumably evolve up to human intelligence.

(11) A program is simply a string of 0's and 1's. For a program to be, generated randomly it means that whenever the computer asks for another bit of input, we simply toss a fair coin and give the computer a 1, say, if the coin comes up heads and a 0 if the coin reads tails. This argument may not apply to genetic algorithms which may meet the criteria for evolution and randomness. Casti, pg. 171.

(12) Bennett, Deborah J. "Randomness", Harvard University Press, 1998. pg. 142.

(13) Bennett, pg. 143-148.

(14) Russell, Stuart J. and Norvig, Peter. "Artificial Intelligence: A Modern Approach". Prentice Hall, 1995.

(15) Casti, pg. 145-146. A continuum has a cardinality (or size of infinity) that is greater than that of the standard countable infinity. For more on Cantorís Diagonal Argument, see Penrose, Ch. 2.

(16) Physicists like Penrose would conclude that this randomness is actualized through inherently random quantum mechanical processes in the brain. Penrose, Roger. "The Emperor's New Mind." Oxford University Press, 1989.,

(17) To test human ability as a random number generator, we could have a human subject construct a random sequence by speaking words in a non sequitor fashion, where some external source has already decided how to translate those words into binary digits, where the method is unknown to the subject. After generating a binary sequence this way, we could compare results to approximate random number generators, and apply statistical tests to see which sequence is "more random". Such a process would be largely intractable in practice, and as such it is better to view it theoretically as a testament to the differences we may have, in principle.

(18) In this context, unconscious processes are simply those that we have no means of accessing by direct introspection. For example, although we could read about kidneys in a biology textbook, there is no way we can be consciously aware of the biological details, which are ordered and algorithmic processes. This usage of the word "unconscious" is to be differentiated from the "unconscious mind" of psychologists like Freud and Jung, since we can access this part of our mind through introspection via dreams, hypnosis, or drugs. In any case, the Freudian playground of dreams and neuroses is most certainly non-algorithmic and largely dependent on randomness.

(19) Penrose, Ch. 4.

(20) Casti, pg. 167

(21) Casti, pg. 168.

(22) Hofstadter, Douglas. "Godel, Escher, Bach: An Eternal Golden Braid", Basic Books, New York, 1979.

(23) This does not imply that the sentence L is meaningless, only that the question, "Is L true or false?" is ill posed in this context.

(24) In contrast, Turing Machines are, by definition, isomorphic to consistent formal mathematical systems. Casti, pg. 156.

Consistent formal systems could not contain inconsistent statements like L, whose truth and falsity can be derived simultaneously. However, whether we view such paradoxes as inconsistent statements depends on whether language itself can be thought of as a true formal system, and whether it is proper to assign truth value to sentences from the meta-language. It should also be noted that the Godel statement G which says, "Statement G is unprovable within the system", is still consistent since it is merely true, and does not lead to a contradiction like L does.

(25) Casti, pg. 168.

(26) Barrow, John D. "Impossibility", Oxford University Press, 1998. pg. 230-232.

On a side note, if human language is inconsistent, and if mathematical linguistic models such as Noam Chomsky's Transformational Generative Grammar are consistent, then those models are at best only an extremely useful approximation to the human language generator. Chomsky, Noam. "*Recent Contributions to the Theory of Innate Ideas*", J.R Searle, ed. The Philosophy of Language, Oxford University Press, 1971, pp. 121-128.

(27) Barrow, pg. 232.

Graphics created by Tom Yedwab.

Andrew Friedman is a recent Berkeley graduate, completing a degree with a double major in Physics and Astrophysics.

***Please send questions or comments to*:**
andyf365@uclink4.berkeley.edu