To appear in: Epstein, Robert & Peters, Grace (Eds.) The Turing Test Sourcebook: Philosophical and Methodological Issues in the Quest for the Thinking Computer. Kluwer.

The Annotation Game: On Turing (1950) on Computing, Machinery, and Intelligence.

Stevan Harnad

> **I propose to consider the question, "Can machines think?" (Turing 1950)**

Turing starts on an equivocation. We know now that what he will go on to consider is not whether or not machines can think, but whether or not machines can <u>do</u> what thinkers like us can do -- and if so, <u>how</u>. Doing is performance capacity, empirically observable. Thinking (or cognition) is an internal state, its correlates empirically observable as neural activity (if we only knew which neural activity corresponds to thinking!) and its associated quality introspectively observable as our own mental state when we are thinking. Turing's proposal will turn out to have nothing to do with either observing neural states or introspecting mental states, but only with generating performance capacity (intelligence?) indistinguishable from that of thinkers like us.

> **This should begin with definitions of... "machine" and "think"... [A] statistical survey such as a Gallup poll [would be] absurd. Instead of attempting such a definition I shall replace the question by another... in relatively unambiguous words.**

"Machine" will never be adequately defined in Turing's paper, although (what will eventually be known as) the "Turing Machine," the abstract description of a computer, will be. This will introduce a systematic ambiguity between a real physical system, doing something in the world, and another physical system, a computer, simulating the first system formally, but not actually doing what it does: An example would be the difference between a real airplane -- a machine, flying in the real world -- and a computer simulation of an airplane, not really flying, but doing something formally equivalent to flying, in a (likewise simulated) "virtual world."

A reasonable definition of "machine," rather than "Turing Machine," might be *any dynamical, causal system*. That makes the universe a machine, a molecule a machine, and also waterfalls, toasters, oysters and human beings. Whether or not a machine is man-made is obviously irrelevant. The only relevant property is that it is "mechanical" -- i.e., behaves in accordance with the cause-effect laws of physics (Harnad 2003).

"Think" will never be defined by Turing at all; it will be replaced by an operational definition to the effect that "thinking is as thinking does." This is fine, for thinking (cognition, intelligence) cannot be defined in advance of knowing how thinking systems do it, and we don't yet know how. But we do know that we thinkers do it, whatever it is, when we think; and we know when we are doing it (by introspection). So thinking, a form of consciousness, is already ostensively defined, by just pointing to that experience we all have and know.

Taking a statistical survey like a Gallup Poll instead, to find out people's opinions of what thinking is would indeed be a waste of time, as Turing points out -- but then later in the paper he needlessly introduces the equivalent of a statistical survey as his criterion for having passed his Turing Test!

> **The new form of the problem can be described in terms of a game which we call the 'imitation game."**

Another unfortunate terminological choice: "Game" implies caprice or trickery, whereas Turing in fact means serious empirical business. The game is science (the future science of cognition -- actually a branch of reverse bioengineering; Harnad 1994a). And "imitation" has connotations of fakery or deception too, whereas what Turing will be proposing is a rigorous empirical methodology for testing theories of human cognitive performance capacity (and thereby also theories of the thinking that presumably engenders that capacity). Calling this an "imitation game" (instead of a methodology for reverse-engineering human cognitive performance capacity) has invited generations of needless misunderstanding (Harnad 1992).

> **The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman.**

The man/woman test is an intuitive "preparation" for the gist of what will eventually be the Turing Test, namely, an empirical test of performance capacity. For this, it is first necessary that all non-performance data be excluded (hence the candidates are out of sight). This sets the stage for what will be Turing's real object of comparison, which is a thinking human being versus a (nonthinking) machine, a comparison that is to be unbiased by appearance.

Turing's criteria, as we all know by now, will turn out to be two (though they are often confused or conflated): (1) Do the two candidates have identical performance capacity? (2) Is there any way we can distinguish them, based only on their performance capacity, so as to be able to detect that one is a thinking human being and the other is just a machine? The first is an empirical criterion: Can they both do the same things? The second is an intuitive criterion, drawing on what decades later came to be called our human "mind-reading" capacities (Frith & Frith 1999): Is there anything about the way the candidates go about doing what they can both do that cues me to the fact that one of them is just a machine?

Turing introduces all of this in the form of a party game, rather like 20-Questions. He never explicitly debriefs the reader to the effect that what is really at issue is no less than the game of life itself, and that the "interrogator" is actually the scientist for question (1), and, for question (2), any of the rest of us, in every one of our daily interactions with one another. The unfortunate party-game metaphor again gave rise to needless cycles of misunderstanding in later writing and thinking about the Turing Test.

> **In order that tones of voice may not help the interrogator, the answers should be written, or better still, typewritten.**

This restriction of the test exclusively to what we would today call email interactions is, as noted, a reasonable way of preparing us for its eventual focus on performance capacity alone, rather than appearance, but it does have the unintended further effect of ruling out all direct testing of performance capacities other than verbal ones; and that is potentially a much more serious equivocation, to which we will return. For now, we should bear in mind only that if the criterion is to be Turing-indistinguishable performance-capacity, we can all do a lot more than just email!

> **We now ask the question, "What will happen when a machine takes the part of A in this game?" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, "Can machines think?"**

Here, with a little imagination, we can already scale up to the full Turing Test, but again we are faced with a needless and potentially misleading distraction: Surely the goal is not merely to design a machine that people mistake for a human being statistically as often as not! That would reduce the Turing Test to the Gallup Poll that Turing rightly rejected in raising the question of what "thinking" is in the first place! No, if Turing's indistinguishability-criterion is to have any empirical substance, the performance of the machine must be totally indistinguishable from that of a human being -- to anyone and everyone, for a lifetime (Harnad 1989).

> **The new problem has the advantage of drawing a fairly sharp line between the physical and the intellectual capacities of a man.**

It would have had that advantage, if the line had only been drawn between appearance and performance, or between structure and function. But if the line is instead between verbal and nonverbal performance capacities then it is a very arbitrary line indeed, and a very hard one to defend. As there is no explicit or even inferrable defense of this arbitrary line in any of Turing's paper (nor of an equally arbitrary line between those of our "physical capacities" that do and do not depend on our "intellectual capacities"), I will take it that Turing simply did not think this through. Had he done so, the line would have been drawn between the candidate's physical appearance and structure on the one hand, and its performance capacities, both verbal and nonverbal, on the other. Just as (in the game) the difference, if any, between the man and the woman must be detected from what they <u>do</u>, and not what they look like, so the difference, if any, between human and machine must be detected from what they do, and not what they look like. This would leave the door open for the robotic version of the Turing Test that we will discuss later, and not just for the email version.

But before a reader calls my own dividing line between structure and function just as arbitrary, let me quickly agree that Turing has in fact introduced a hierarchy of Turing Tests here, but not an infinity of them (Harnad 2000). The relevant levels of this hierarchy will turn out to be only the following 5:

> **t0**: *Local indistinguishability in the capacity to perform some arbitrary task, such as chess*. t0 is not

really a Turing Test at all, because it is so obviously subtotal; hence the machine candidate is easily distinguished from a human being by seeing whether it can do anything else, other than play chess. If it can't, it fails the test.

T2: *Total indistinguishability in email (verbal) performance capacity*. This seems like a self-contained performance module, for one can talk about anything and everything, and language has the same kind of universality that computers (Turing Machines) turned out to have. T2 even subsumes chess-playing. But does it subsume star-gazing, or even food-foraging? Can the machine go and see and then tell me whether the moon is visible tonight and can it go and unearth truffles and then let me know how it went about it? These are things that a machine with email capacity alone cannot do, yet every (normal) human being can.

T3: *Total indistinguishablity in robotic (sensorimotor) performance capacity*. This subsumes T2, and is (I will argue) the level of Test that Turing really intended (or should have!).

T4: *Total indistinguishability in external performance capacity as well as in internal structure/function*. This subsumes T3 and adds all data that a neuroscientist might study. This is no longer strictly a Turing Test, because it goes beyond performance data, but it correctly embeds the Turing Hierarchy in a larger empirical hierarchy. Moreover, the boundary between T3 and T4 really is fuzzy: Is blushing T3 or T4?

> **T5**: *Total indistinguishability in physical structure/function*. This subsumes T4 and rules out any functionally equivalent but synthetic nervous systems: The T5 candidate must be indistinguishable from other human beings right down to the last molecule.

> **No engineer or chemist claims to be able to produce a material which is indistinguishable from the human skin... [There would be] little point in trying to make a "thinking machine" more human by dressing it up in such artificial flesh.**

Here Turing correctly rejects T5 and T4 -- but certainly not T3.

> **The form in which we have set the problem reflects this fact in the condition which prevents the interrogator from seeing or touching the other competitors, or hearing their voices.**

Yes, but using T2 as the example has inadevertently given the impression that T3 is excluded too: not only can we not see or touch the candidate, but the candidate cannot see or touch anything either -- or do anything other than compute and email.

> **The question and answer method seems to be suitable for introducing almost any one of the fields of human endeavour that we wish to include.**

This correctly reflects the universal power of natural language (to say and describe anything in words). But "almost" does not fit the Turing criterion of identical performance capacity.

> **We do not wish to penalise the machine for its inability to shine in beauty competitions**

This is the valid exclusion of appearance (moreover, most of us could not shine in beauty competitions either).

> **nor to penalise a man for losing in a race against an aeroplane**

Most of us could not beat Deep Blue at chess, nor even attain ordinary grandmaster level. It is only generic human capacities that are at issue, not those of any specific individual. On the other hand, just about all of us can walk and run. And even if we are handicapped (an anomalous case, and hardly the one on which to build one's attempts to generate positive performance capacity), we all have <u>some</u> sensorimotor capacity. (Neither Helen Keller nor Stephen Hawking is a disembodied email-only module.)

> **The conditions of our game make these disabilities irrelevant**

Disabilities and appearance are indeed irrelevant. But nonverbal performance capacities certainly are not. Indeed, our verbal abilities may well be *grounded* in our nonverbal abilities (Harnad 1990; Cangelosi & Harnad

2001; Kaplan & Steels 1999). (Actually, by "disability," Turing means non-ability, i.e., absence of an ability; he does not really mean being disabled in the sense of being physically handicapped, although he does mention Helen Keller later.)

> **the interrogator cannot demand practical demonstrations**

This would definitely be a fatal flaw in the Turing Test if Turing had meant it to exclude T3 -- but I doubt he meant that. He was just arguing that it is performance capacity that is decisive (for the empirical problem that future cognitive science would eventually address), not something else that might depend on irrelevant features of structure or appearance. He merely used verbal performance as his intuition-priming example, without meaning to imply that all "thinking" is verbal and only verbal performance capacity is relevant.

> **The question... will not be quite definite until we have specified what we mean by the word "machine." It is natural that we should wish to permit every kind of engineering technique to be used in our machines.**

This passage (soon to be contradicted in the subsequent text!) implies that Turing did not mean only computers: that any dynamical system we build is eligible (as long as it delivers the performance capacity). But we do have to build it, or at least have a full causal understanding of how it works. A cloned human being cannot be entered as the machine candidate (because we didn't build it and hence don't know how it works), even though we are all "machines" in the sense of being causal systems (Harnad 2000, 2003).

> **We also wish to allow the possibility that an engineer or team of engineers may construct a machine which works, but whose manner of operation cannot be satisfactorily described by its constructors because they have applied a method which is largely experimental.**

Here is the beginning of the difference between the field of artificial intelligence (AI), whose goal is merely to generate a useful performance tool, and cognitive modeling (CM), whose goal is to explain how human cognition is generated. A device we built but without knowing how it works would suffice for AI but not for CM.

> **Finally, we wish to exclude from the machines men born in the usual manner.**

This does not, of course, imply that we are not machines, but only that the Turing Test is about finding out what kind of machine we are, by designing a machine that can generate our performance capacity, but by causal/functional means that we understand, because we designed them.

> **[I]t is probably possible to rear a complete individual from a single cell of the skin (say) of a man... but we would not be inclined to regard it as a case of "constructing a thinking machine."**

This is because we want to explain thinking capacity, not merely duplicate it.
http://www.ecs.soton.ac.uk/~harnad/Hypermail/Foundations.Cogniti

> **This prompts us to abandon the requirement that every kind of technique should be permitted. We [accordingly] only permit digital computers to take part in our game.**

This is where Turing contradicts what he said earlier, withdrawing the eligibility of all engineering systems but one, thereby introducing another arbitrary restriction -- one that would again rule out T3. Turing earlier said (correctly) that any engineering device ought to be eligible. Now he says it can only be a computer. His motivation is partly of course the fact that the computer (Turing Machine) has turned out to be universal, in that it can simulate any other kind of machine. But here we are squarely in the T2/T3 equivocation, for a simulated robot in a virtual world is neither a real robot, nor can it be given a real robotic Turing Test, in the real world. Both T2 and T3 are tests conducted in the real world. But an email interaction with a virtual robot in a virtual world would be T2, not T3.

To put it another way: With the Turing Test we have accepted, with Turing, that thinking is as thinking does. But we know that thinkers can and do do more than just talk. And it remains what thinkers can do that our candidate must likewise be able to do, not just what they can do verbally. Hence just as flying is something that only a real plane can do, and not a computer-simulated virtual plane, be it ever so Turing-equivalent to the real plane -- so passing T3 is something only a real robot can do, not a simulated robot tested by T2, be it ever so

Turing-equivalent to the real robot. (I also assume it is clear that Turing Testing is testing in the real world: A virtual-reality simulation [VR] would be no kind of a Turing Test; it would merely be fooling our senses in the VR chamber rather than testing the candidate's real performance capacity in the real world.)

So the restriction to computer simulation, though perhaps useful for planning, designing and even pre-testing the T3 robot, is merely a practical methodological strategy. In principle, any engineered device should be eligible; and it must be able to deliver T3 performance, not just T2.

It is of interest that contemporary cognitive robotics has not gotten as much mileage out of computer-simulation and virtual-worlds as might have been expected, despite the universality of computation. "Embodiment" and "situatedness" (in the real world) have turned out to be important ingredients in empirical robotics (Brooks 2002, Kaplan & Steels 1999), with the watchword being that the real world is better used as its own model (rather than roboticists' having to simulate, hence second-guess in advance, not only the robot, but the world too).

The impossibility of second-guessing the robot's every potential "move" in advance, in response to every possible real-world contingency also points to a latent (and I think fatal) flaw in T2 itself: Would it not be a dead give-away if one's email T2 pen-pal proved incapable of ever commenting on the analog family photos we kept inserting with our text? (If he can process the images, he's not just a computer but at least a computer plus A/D peripheral sensors, already violating Turing's arbitrary restriction to computers alone). Or if one's pen-pal were totally ignorant of contemporaneous real-world events, apart from those we describe in our letters? Wouldn't even its verbal performance break down if we questioned it too closely about the qualitative and practical details of sensorimotor experience? Could all of that really be second-guessed purely verbally in advance?

> **This restriction [to computers] appears at first sight to be a very drastic one. I shall attempt to show that it is not so in reality. To do this necessitates a short account of the nature and properties of these computers.**

The account of computers that follows is useful and of course correct, but it does not do anything at all to justify restricting the TT to candidates that are computers. Hence this arbitrary restriction is best ignored.

> **It may also be said that this identification of machines with digital computers, like our criterion for "thinking," will only be unsatisfactory if (contrary to my belief), it turns out that digital computers are unable to give a good showing in the game.**

This is the "game" equivocation again. It is not doubted that computers will give a good showing, in the Gallup Poll sense. But empirical science is not just about a good showing: An experiment must not just fool most of the experimentalists most of the time! If the performance-capacity of the machine must be indistinguishable from that of the human being, it must be totally indististinguishable, not just indististinguishable more often than not. Moreover, some of the problems that I have raised for T2 -- the kinds of verbal exchanges that draw heavily on sensorimotor experience -- are not even likely to give a good showing, if the candidate is a digital computer only, regardless of how rich a data-base it is given in advance.

> **[D]igital computers... carry out any operations which could be done by a human computer... following fixed rules...**

This goes on to describe what has since bcome the standard definition of computers as rule-based symbol-manipulating devices (Turing Machines).

> **An interesting variant on the idea of a digital computer is a "digital computer with a random element"... Sometimes such a machine is described as having free will (though I would not use this phrase myself)**

Nor would I. But surely an even more important feature for a Turing Test candidate than a random element or statistical functions would be autonomy in the world -- which is something a T3 robot has a good deal more of than a T2 pen-pal. The ontic side of free will -- namely, whether we ourselves, real human beings, actually have free will -- rather exceeds the scope of Turing's paper Harnad 1982b). So too does the question of whether a TT-passing machine would have any feelings at all (whether free or otherwise)

(Harnad 1995). What is clear, though, is that computational rules are not the only ways to "bind" and determine performance: ordinary physical causality can do so too.

> **It will seem that given the initial state of the machine and the input signals it is always possible to predict all future states. This is reminiscent of Laplace's view that from the complete state of the universe at one moment of time, as described by the positions and velocities of all particles, it should be possible to predict all future states.**

The points about determinism are probably red herrings. The only relevant property is performance capacity. Whether either the human or the machine are completely predictable is irrelevant. (Both many-body physics and complexity theory suggest that neither causal determinacy nor rulefulness guarantee predictability in practise -- and this is without even invoking the arcana of quantum theory.)

> **Provided it could be carried out sufficiently quickly the digital computer could mimic the behavior of any discrete-state machine... they are universal machines.... [Hence] considerations of speed apart, it is unnecessary to design various new machines to do various computing processes. They can all be done with one digital computer, suitably programmed for each case... [A]s a consequence of this, all digital computers are in a sense equivalent.**

All true, but all irrelevant to the question of whether a digital computer alone could pass T2, let alone T3. The fact that eyes and legs can be simulated by a computer does not mean a computer can see or walk (even when it is simulating seeing and walking). So much for T3. But even just for T2, the question is whether simulations alone can give the T2 candidate the capacity to verbalize and converse about the real world indistinguishably from a T3 candidate with autonomous sensorimotor experience in the real world.

(I think yet another piece of unnoticed equivocation by Turing -- and many others -- arises from the fact that thinking is not observable: That unobservability helps us imagine that computers think. But even without having to invoke the other-minds problem (Harnad 1991), one needs to remind oneself that a universal computer is only formally universal: It can describe just about any physical system, and simulate it in symbolic code, but in doing so, it does not capture all of its properties: Exactly as a computer-simulated airplane cannot really do what a plane plane does (i.e., fly in the real-world), a computer-simulated robot cannot really do what a real robot does (act in the real-world) -- hence there is no reason to believe it is really thinking either. A real robot may not really be thinking either, but that does require invoking the other-minds problem, whereas the virtual robot is already disqualified for exactly the same reason as the virtual plane: both fail to meet the TT criterion itself, which is real performance capacity, not merely something formally equivalent to it!)

> **I believe that in about fifty years' time it will be possible, to programme computers... [to] 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.**

No doubt this party-game/Gallup-Poll criterion can be met by today's computer programs -- but that remains as meaningless a demographic fact today as it was when predicted 50 years ago: Like any other science, cognitive science is not the art of fooling most of the people for some or most of the time! The candidate must really have the generic performance capacity of a real human being -- capacity that is totally indistinguishable from that of a real human being to any real human being (for a lifetime, if need be!). No tricks: real performance capacity.

> **The original question, "Can machines think?" I believe to be too meaningless to deserve discussion.**

It is not meaningless, it is merely undecidable: What we mean by "think" is, on the one hand, what thinking creatures can do and how they can do it, and, on the other hand, what it feels-like to think. What thinkers can do is captured by the TT. A theory of how they do it is provided by how our man-made machine does it. (If there are several different successful machines, it's a matter of normal inference-to-the-best-theory.) So far, nothing meaningless. Now we ask: Do the successful candidates really feel, as we do when we think? This question is not meaningless, it is merely unanswerable -- in any other way than by being the candidate. It is the familar old other-minds problem (Harnad 1991).

> **Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.**

Yes, but only at a cost of demoting "thinking" to meaning only "information processing" rather than what you or I do when we think, and what that feels-like.

> **The popular view that scientists proceed inexorably from well-established fact to well-established fact, never being influenced by any improved conjecture, is quite mistaken. Provided it is made clear which are proved facts and which are conjectures, no harm can result. Conjectures are of great importance since they suggest useful lines of research.**

This is confused. Yes, science proceeds by a series of better approximations, from empirical theory to theory. But the theory here would be the actual design of a successful TT candidate, not the conjecture that computation (or anything else) will eventually do the trick. Turing is confusing formal conjectures (such as that the Turing Machine and its equivalents capture all future notions and instances of what we mean by "computation" -- the so-called "Church/Turing Thesis") and empirical hypotheses, such as that thinking is just computation. Surely the Turing Test is not a license for saying that we are explaining thinking better and better as our candidates fool more and more people longer and longer. On the other hand, something else that sounds superficiallly similar to this (but happens to be correct) could be said about scaling up to the TT empirically by designing a candidate that can do more and more of what we can do. And Turing Testing certainly provides a methdology for such cumulative theory-building and theory-testing in cognitive science.

> **The Theological Objection: Thinking is a function of man's immortal soul.**

The real theological objection is not so much that the soul is immortal but that it is immaterial. This view also has non-theological support from the mind/body problem: No one -- theologian, philosopher or scientist -- has even the faintest hint of an idea of how mental states can be material states (or, as I prefer to put it, how functional states can be felt states). This problem has been dubbed "hard" (Chalmers in Shear 1997). It may be even worse: it may be insoluble (Harnad 2001). But this is no objection to Turing Testing which, even if it will not explain how thinkers can feel, does explain how they can do what they can do.

> **[T] here is a greater difference, to my mind, between the typical animate and the inanimate than there is between man and the other animals.**

Yes, and this is why the other-minds problem comes into its own in doing Turing-Testing of machines rather than in doing mind-reading of our own species and other animals. ("Animate" is a weasel-word, though, for vitalists are probably also animists; Harnad 1994a.)

> **The Mathematical Objection: Godel's theorem[:] [A]lthough 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.**

Godel's theorem shows that there are statements in arithmetic that are true, and we know are true, but their truth cannot be computed. Some have interpreted this as implying that "knowing" (which is just a species of "thinking") cannot be just computation. Turing replies that maybe the human mind has similar limits, but it seems to me it would have been enough to point out that "knowing" is not the same as "proving." Godel shows the truth is unprovable, not that it is unknowable. (There are far better reasons for believing that thinking is not computation.)

> **The Argument from Consciousness:**
>
> **Jefferson (1949): "Not until a machine can [do X] because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain"**

This standard argument against the Turing Test (repeated countless times in almost exactly the same way until the present day) is merely a restatement of the other-minds problem: There is no way to know whether either humans or machines do what they does because they feel like it -- or whether they feel anything at all, for that matter. But there is a lot to be known from identifying what can and cannot generate the capacity to do what humans can do. (The limits of symbol-manipulation [computation] are another matter, and one that can be settled empirically, based on what sorts of machine can and cannot pass the TT; Harnad 2003.)

> **According to the most extreme form of this view the only way by which one could be sure that machine thinks is to be the machine and to feel oneself thinking... [This] is in fact the solipsist point of view. It may be the most logical view to hold but it makes communication of ideas**

> **difficult.**

Turing is dead wrong here. This is not solipsism (i.e., not the belief that only I exist and all else is my dream). It is merely the other-minds problem (Harnad 1991); and it is correct, but irrelevant -- or rather put into perspective by the Turing Test: There is no one else we can know has a mind but our own private selves, yet we are not worried about the minds of our fellow-human beings, because they behave just like us and we know how to mind-read their behavior. By the same token, we have no more or less reason to worry about the minds of anything else that behaves just like us -- so much so that we can't tell them apart from other human beings. Nor is it relevant what stuff they are made out of, since our successful mind-reading of other human beings has nothing to do with what stuff they are made out of either. It is based only on what they do.

> **Arguments from Various [In]abilities:**
> **... "I grant you that you can make machines do all the things you have mentioned but you will never be able to make one to do X" : [e.g] Be kind, resourceful, beautiful, friendly, have initiative, have a sense of humour, tell right from wrong, make mistakes, fall in love, enjoy strawberries and cream, make some one fall in love with it, learn from experience, use words properly, be the subject of its own thought, have as much diversity of behaviour as a man, do something really new.**

Turing rightly dismisses this sort of scepticism (which I've dubbed "Granny Objections" http://www.ecs.soton.ac.uk/~harnad/CM302/Granny/sld001.htm) by pointing out that these are empirical questions about what computers (and other kinds of machines) will eventually be shown to be able to do. The performance items on the list, that is. The mental states (feelings), on the other hand, are moot, because of the other-minds problem.

> **(6) Lady Lovelace's Objection:**
> **"The Analytical Engine has no pretensions to *originate* anything. It can do *whatever we know how to order it* to perform"... a machine can "never do anything really new."**

This is one of the many Granny objections. The correct reply is that (i) all causal systems are describable by formal rules (this is the equivalent of the Church/Turing Thesis), including ourselves; (ii) we know from complexity theory as well as statistical mechanics that the fact that a system's performance is governed by rules does not mean we can predict everything it does; (iii) it is not clear that anyone or anything has "originated" anything new since the Big Bang.

> **The view that machines cannot give rise to surprises is due, I believe, to a fallacy to which philosophers and mathematicians are particularly subject. This is the assumption that as soon as a fact is presented to a mind all consequences of that fact spring into the mind simultaneously with it. It is a very useful assumption under many circumstances, but one too easily forgets that it is false. A natural consequence of doing so is that one then assumes that there is no virtue in the**

> **mere working out of consequences from data and general principles.**

Turing is quite right to point out that knowing something is true does not mean knowing everything it entails; this is especially true of mathematical conjectures, theorems, and axioms.

But I think Lady Lovelace's preoccupation with freedom from rules and novelty is even more superficial than this. It takes our introspective ignorance about the causal basis of our performance capacities at face-value, as if that ignorance demonstrated that our capacities are actually sui generis acts of our psychokinetic will -- rather than being merely the empirical evidence of our functional ignorance, for future reverse-engineering (cognitive science) to remedy.

> **Argument from Continuity in the Nervous System: It may be argued that... one cannot expect to be able to mimic the behaviour of the nervous system with a discrete-state system.**

According to the Church/Turing Thesis, there is almost nothing that a computer cannot simulate, to as close an approximation as desired, including the brain. But, as noted, there is no reason computers should be the only machines eligible for Turing Testing. Robots can have analog components too. Any dynamical causal system is eligible, as long as it delivers the peformance capacity.

> **The Argument from Informality of Behaviour: It is not possible to produce a set of rules purporting to describe what a man should do in**

> **every conceivable set of circumstances.**

First, the successful TT candidate need not be just computational (rule-based); all the arguments for T3 robots and their need of real-world sensorimotor capacities, mechanisms and experience suggest that more is required in a successful candidate than just computation. The impossibility of second-guessing a set of rules that predicts every contingency in advance is probably also behind the so-called "Frame Problem" in Artificial Intelligence (Harnad 1993). But it will still be true, because of the Church-Turing Thesis, that the successful hybrid computational/dynamic T3 robot still be computer-simulable in principle -- a virtual robot in a virtual world. So the rule-based system can <u>describe</u> what a T3 robot would do under all contingencies; that simulation would simply not <u>be</u> a T3 robot, any more than its virtual world would be the real world.

> **Learning Machines**

Turing successfully anticipates machine learning, developmental modeling and evolutionary modeling in this prescient section.

> **The Argument from Extrasensory Perception:... telepathy, clairvoyance, precognition and psychokinesis.... [T]he statistical evidence, at least for telepathy, is overwhelming.**

It is a pity that at the end Turing reveals his credulousness about these dubious phenomena, for if psychokinesis (mind over matter) were genuinely possible, then ordinary matter/energy engineering would not be enough to generate a thinking mind; and if telepathy (true mind-reading) were genuinely possible, then that would definitely trump the Turing Test.

REFERENCES

Brooks, R. A., (2002) Flesh and Machines, Pantheon Books.

Cangelosi, A. & Harnad, S. (2001) The Adaptive Advantage of Symbolic Theft Over Sensorimotor Toil: Grounding Language in Perceptual Categories. *Evolution of Communication*. 4(1) 117-142 http://cogprints.soton.ac.uk/documents/disk0/00/00/20/36/

Frith Christopher D. & Frith, Uta (1999) Interacting minds Ð a biological basis. Science 286: 1692Ð1695. http://pubpages.unh.edu/~jel/seminar/Frith_mind.pdf

Harnad, S. (1982a) Neoconstructivism: A Unifying Constraint for the Cognitive Sciences, In: Language, mind and brain (T. Simon & R. Scholes, eds., Hillsdale NJ: Erlbaum), 1 - 11. http://cogprints.soton.ac.uk/documents/disk0/00/00/06/62/

Harnad, S. (1982b) Consciousness: An afterthought. Cognition and Brain Theory 5: 29 - 47. http://cogprints.soton.ac.uk/documents/disk0/00/00/15/70/

Harnad, S. (1989) Minds, Machines and Searle. Journal of Theoretical and Experimental Artificial Intelligence 1: 5-25. http://cogprints.soton.ac.uk/documents/disk0/00/00/15/73/

Harnad, S. (1990) The Symbol Grounding Problem Physica D 42: 335-346. http://cogprints.soton.ac.uk/documents/disk0/00/00/06/15/

Harnad, S. (1991) "Other Bodies, Other Minds: A Machine Incarnation of an Old Philosophical Problem"Minds and Machines 1: 43-54. http://cogprints.soton.ac.uk/documents/disk0/00/00/15/78/

Harnad, S. (1992) The Turing Test Is Not A Trick: Turing Indistinguishability Is A Scientific Criterion. SIGART Bulletin 3(4) (October 1992) pp. 9 - 10. http://cogprints.soton.ac.uk/documents/disk0/00/00/15/84/

Harnad, Stevan (1993) Problems, Problems: the Frame Problem as a Symptom of the Symbol Grounding Problem, Psycoloquy: 4,#34 Frame Problem (11) http://psycprints.ecs.soton.ac.uk/archive/00000328/

Harnad, S. (1993) Grounding Symbols in the Analog World with Neural Nets. Think 2(1) 12-78. http://psycprints.ecs.soton.ac.uk/archive/00000163/

Harnad, S. (1994a) Levels of Functional Equivalence in Reverse Bioengineering: The Darwinian Turing Test for Artificial Life. Artificial Life 1(3): 293-301. Reprinted in: C.G. Langton (Ed.). Artificial Life: An Overview. MIT Press 1995. http://cogprints.soton.ac.uk/documents/disk0/00/00/15/91/-

Harnad, S. (1994b) Computation Is Just Interpretable Symbol Manipulation: Cognition Isn't. Special Issue on "What Is Computation" Minds and Machines 4:379-390 http://cogprints.soton.ac.uk/documents/disk0/00/00/15/92/

Harnad, Stevan (1995) "Why and How We Are Not Zombies. Journal of Consciousness Studies1:164-167.

http://cogprints.soton.ac.uk/documents/disk0/00/00/16/01/

Harnad, S. (2000) Minds, Machines, and Turing: The Indistinguishability of Indistinguishables. Journal of Logic, Language, and Information 9(4): 425-445. (special issue on "Alan Turing and Artificial Intelligence") http://cogprints.soton.ac.uk/documents/disk0/00/00/16/16/

Harnad, S. (2001) No Easy Way Out. The Sciences 41(2) 36-42. http://cogprints.soton.ac.uk/documents/disk0/00/00/16/24/

Harnad, S. (2002a) Turing Indistinguishability and the Blind Watchmaker. In: J. Fetzer (ed.) Evolving Consciousness Amsterdam: John Benjamins. Pp 3-18. http://cogprints.soton.ac.uk/documents/disk0/00/00/16/15/

Harnad, S. (2002b) Darwin, Skinner, Turing and the Mind. Inaugural Address. Hungarian Academy of Science. http://www.ecs.soton.ac.uk/~harnad/Temp/darwin.htm

Harnad, S. (2003) Can a Machine Be Conscious? How? Journal of Consciousness Studies. http://www.ecs.soton.ac.uk/~harnad/Temp/machine.htm

Shear, J. (Ed.) (1997) Explaining Consciousness: The Hard Problem. MIT/Bradford http://www.u.arizona.edu/~chalmers/book/hp-contents.html

Steels, L. and Kaplan, F. (1999) Situated grounded word semantics. In Dean, T., editor, Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence IJCAI'99, pages 862-867, San Francisco, CA., 1999. Morgan Kaufmann Publishers. http://arti.vub.ac.be/steels/ijcai99.pdf