1) The question "can machines think?" misguides our understanding, as we have imprecise notions of "machine" and "thought"; the question should be reframed.

Turing proposes to reframe the question in a new form, as a game, in which an interrogator in a different room attempts to determine the gender of two entities – one "man" and one "woman" – purely on the basis of their answers to questions that it puts to them.

One entity in the room aims to trick the interrogator, whereas the other aims to gain trust.

Questions and answers are written, or communicated in such a way that the manner of articulation of the words cannot convey meaningful information about the potentially machinic nature of their source.

The question "can machines think?" is thus reframed as "can a machine take the place of the decieving entity in this game such that the interrogator is deceived as often as when both entities are human?"

2) Turing is interested both in the direct answer to this question, and to whether it is a question that yields interesting research.

Turing thinks it is good that the form of this question draws a distinct line between the <u>physical</u> and <u>intellectual</u> capacities of the human.

Time between question and answer is included as a signal to the interrogator.

These conditions allow us to compare the human and the machine in terms of intellectual capability alone, and not in other domains of capability. Is such a demarcation between intellectual and other capabilities really proper to the activity of "thinking"?

Perhaps the game's conditions even too strongly disadvantage the machine, i.e. it is testing something "too human."

Another potential objection, that the machine's optimal strategy could involve doubling down on its inhumanity, is dismissed as unlikely.

3) Turing makes the idea of a "machine" more precise by excluding from it something which is human, and moreover specifying that it should refer to what we can more precisely call a <u>digital computer</u>. Turing's gendered association of human with man is problematic.

The reduction of the category of "machine" to mean "digital computer" is not as drastic as it may first appear.

If digital computers can be good at the imitation game, it will retroactively justify this definitional narrowing. This line of argumentation seems dubious...

By "digital computer", Turing means a specific concept, not a specific material artifact.

4) The digital computer is an entity that follows fixed rules, with no deviation whatsoever

A computer has three components: 1) a store, 2) an executive unit, 3) a control.

A <u>store</u> is the substance on/in which records can be made.

An <u>executive unit</u> is the motive force that carries out operations.

A <u>control</u> is the set of fixed instructions the executive unit carries out to modify the store

The store can be considered as broken up into "packets".

Turing gives an example of a computer at work in multiplication

Instructions are followed sequentially, but can 'jump' back to positions previously executed so that repetition is possible. The example with "Mother" and "Tommy" is again interestingly gendered.

The idea of a digital computer can be built in material reality.

Constructing instruction tables for a digital computer is called <u>programming</u>.

The definition of a computer (having "fixed instructions") does not preclude random elements, such as arbitrary jumps. Turing's comment about the difficulty of observing pure randomness as opposed to pseudorandomness anticipates the decades of mathematical thought that will develop on this point, in cryptography and complexity theory.

Though computers can theoretically have an infinite store, in practice there is always a material limit.

The <u>idea</u> of a computer has existed at least since Babbage, in the early 19th century.

Nothing about the idea of a computer requires that it be electrical; this is simply one (very effective) material manifestation of it.

5) A computer is a subcategory of a <u>discrete state machine</u>, which is a theoretical construct rather than a material reality.

Each state of the machine can be fully encoded into a table. Time, in a sense, is removed as a concern; everything is instead represented spatially, as a table.

We can only reasonably consider something a discrete state machine if the progression of its states is predictable, i.e. fully deterministic. There is a tension here with Turing's admission of random elements a few paragraphs before; the reference to Laplace is interesting when thinking about the history of philosophy of mechanism/determinism.

The number of states, and symbols that represent states, must also be materially constrained.

A digital computer (in practice) can thus be considered a materially constrained discrete state machine.

As a digital computer can mimic any discrete state machine, we can call it a <u>universal</u> machine.

The paper's opening question is further refined to: can a discrete state machine effectively play the imitation game?

6) Let's now debate that question. Simply an introduction to the new section, but a lot of unnecessary words to do it.

Turing's conjecture: in the year 2000, an interrogator will only be right 70% of the time after 5 minutes of questioning. Did anyone do this experiment in 2000? Is Turing's prediction of a capacity of 10^9 accurate? (I suspect it is off by several orders of magnitude, i.e. 2000 computers had much more storage.)

Considers the <u>Theological Objection</u>, that thinking is a function of "man's immortal soul".

Gloriously lays aside his obvious atheism and answers religion on its own terms: if true, God cannot be omnipotent, as he is not "free" to confer a soul on a rock, or an elephant. Though it seems somewhat specious in brief, Turing's argument here is in fact demonstrating how our own concepts are inadequate to grasp actual reality, a theme that carries through debates in the history of philosophy, from Plato, to Hegel, to Marx, to poststructuralism and psychoanalysis.

Further reinforces his skepticism of the Theological Objection with reference to the Galilean revolution: who are we to claim that we have a clear-headed notion of the soul such that we can decide who/what has one, and who/what doesn't?

Considers the <u>Heads in the Sand Objection</u>, that if machines could "think", the consequences would be too dreadful. Turing slips out of his previous qualification that the question "can machines think?" is too imprecise here.

This is fundamentally a narcissism, often implicit rather than explicit, and common in intellectuals.

Scathingly; no counterargument required.

Considers the <u>Mathematical Objection</u>, which is some application of Gödel's incompleteness theorems (which refer to "logical systems" rather than machines) to refute the capacity of a machine to think. An interesting comparison between machines, discrete state machines, and logical systems drawn here.

Turing agrees that this is a disability of machines, but he sees no grounds for thinking that some form of this disability does not also apply to humans.

The imitation game, Turing thinks, will be reasonable to those with the Mathematical Objection.

Considers <u>The Argument from Consciousness</u>, which claims that "emotions" and "thoughts" must underlie a symbol processing for us to consider it thinking.

Because there seems to be no empirical way to test that such underlying thoughts and emotions exist (beyond being the entity experiencing them oneself), Turing settles on "the polite convention" that everyone thinks.

The imitation game is posited as a reasonable way to confirm whether an entity 'understands' the answers it gives, as we can ask it follow-up questions.

The empirical difference between simulating understanding and actual understanding is difficult (if not impossible) to disentangle.

Thus those with this argument should also accept the imitation game as a reasonable test.

Turing does not want to claim that there only exists that which can be observed, the behaviorist position. This seeming aside is particularly interesting, as it could suggest that Turing is at some odds with the cyberneticians (such as Norbert Weiner) who in the 60s would go on a crusade of behaviorism.

Considers <u>Arguments from Various Disabilities</u>, according to which machines may encroach on certain human activities, but will never replicate some kinds of them.

Turing considers these for the most part intellectually unfounded arguments based on an unconscious assumption of a machine's inanimacy.

Seemingly frivolous activities, such as enjoyment, may be essential to human performance in other activities. Turing mentions race here in relation to the problem of 'enjoyment' – which will be important when thinking about what role psychoanalysis can play in helping to form a reasonable conception of the machine.

Turing introduces the paradox of mistake-making when it comes to machines; if they are to appear as humans they must be capable of certain kinds of error. The relationship between mechanism and error also has a rich philosophical heritage, in particular with respect to Descartes' radical doubt, and his idea of the animal-machine.

Further introduces the paradox of the (self-reflexive) subject when it comes to 'thought' and the machine.

As the storage capacity of machines (digital computers) grows, they will be capable of greater diversity of behavior.

All these debates are disguised forms of the Argument from Consciousness (which Turing already dismissed).

Considers <u>Lady Lovelace's Objection</u>, which is that we must know how to order (program) a machine to perform "thinking" in order for it to be capable of doing it. In other words, we must know more precisely for ourselves what the act of "thinking" entails.

Lovelace's assumptions/intuitions about the capacities of the Analytical Engine should not be generalized to <u>all</u> machines.

A variation of Lovelace's argument is that a machine will never do anything 'new', but do we really have proper resolution on what constitutes something "new"? Turing notes how often machines surprise him, which will resonate with anyone who has done any software programming.

Turing again reduces Lovelace's objection to be another form of the Argument from Consciousness, in that it rests on some slippery notion (such as "new" or "creativity") that is subjectively constituted. In other words, Turing's objections here are that the languages are imprecise.

Turing also puts it to a fallacy in thinking that all consequences of a fact immediately spring to mind, that there is no strong notion of <u>duration</u> in the constitution of consciousness.

Considers the <u>Argument from Continuity in the Nervous System</u>, which is that the nervous system cannot be emulated/simulated by a discrete state machine.

The imitation game constrains the question in such a way that the difference here will not show. Not sure that I fully follow/buy this point.

Considers the <u>Argument from Informality of Behavior</u>, which is that rules of conduct cannot be provided to cover all eventualities. Reminds me of the distinction between morality (rule-based) and ethics.

Rules of conduct ≠ laws of behavior, thus this argument has an "undistributed middle." A great phrase!

The "rules" of conduct may be emergent, not directly cognizable.

Considers the <u>Argument from Extra-Sensory Perception</u>, which addresses phenomena that science and logic cannot (yet) explain. Perhaps incredulously, Turing seems to believe in telepathy.

It is not sufficient to simply pretend that these phenomena do not exist.

E.S.P. between the interrogator and one of the entities could break the conditions of the imitation game. It is easy to get skeptical here, but I think the point Turing makes is actually philosophically critical: what happens to the test set up if externalities enter that defy some of the game's basic assumptions?

To put the argument to bed, Turing makes his room in the imitation game "telepathy-proof."

7) Turing admits to have more in the way of fallacies with the common view that machines are <u>not</u> capable of thought than positive evidence that they actually are. Could we consider Turing's method "immanent" a la Moishe Postone's reading of Marx in this regard?

Developing Lovelace's objection, couldn't we 'inject' an idea into a machine such that it expansively develops a theory, rather than floundering with less than an idea?

Introduces the layers of an onion analogy; are we really just mechanism all the way down?

Re-emphasizes these are not wholly convincing arguments in themselves, just pointers to one.

The most convincing argument Turing offers is to wait for 50 years, then try again. Are we not still waiting, 74 years later?

The problem is fundamentally one of SOFTWARE, i.e. the architecture of abstractions that can give rise to complex behavior. The notion of time is again brought up, and curiously sidelined.

A sufficiently developed machine will likely require <u>education</u> through <u>experience</u>.

Start not with the adult mind, but the (presumably) lower hanging fruit of the child's mind, which is likely to have less architecture of abstraction.

The process of a child-mind plus education resembles <u>evolution</u>.

But Turing expects research to be faster than evolution, as survival of the fittest is a slow heuristic.

The child-machine need not have legs or eyes to be intelligent, as evidenced by Hellen Keller. (!)

Proposes a reward-punishment mechanism to educate such a child-machine. Presaging reinforcement learning's success, in some sense.

Considers that education will be most effective if signal-rewards and punishments are conveyed through "unemotional" channels. One recalls that Turing went to a traditional British all boys public school here...

Introduces the idea that some imperatives should be programmed in, and others learned, and notes that the line between these is not clear-cut.

"Brilliant reasoning" involves making choices about which of multiple valid paths to take. c.f. AlphaGo, and its claim to 'intuition'.

A learning machine must learn to modify certain assumptions, rather than sticking to its facts pretentiously. Interesting analogy to the U.S. constitution...

Intelligent behavior consists in a <u>slight</u> departure from the rigidity of completely disciplined behavior. Presaging deep learning, with its "slight departure" being ReLU.

Further emphasizing this slight departure should have a random element. Again reads as a prescient comment, given gradient descent and suboptimal minima.

Turing believes that machines can thus come to compete with men in all "purely intellectual fields". But how do we demarcate these fields, exactly?