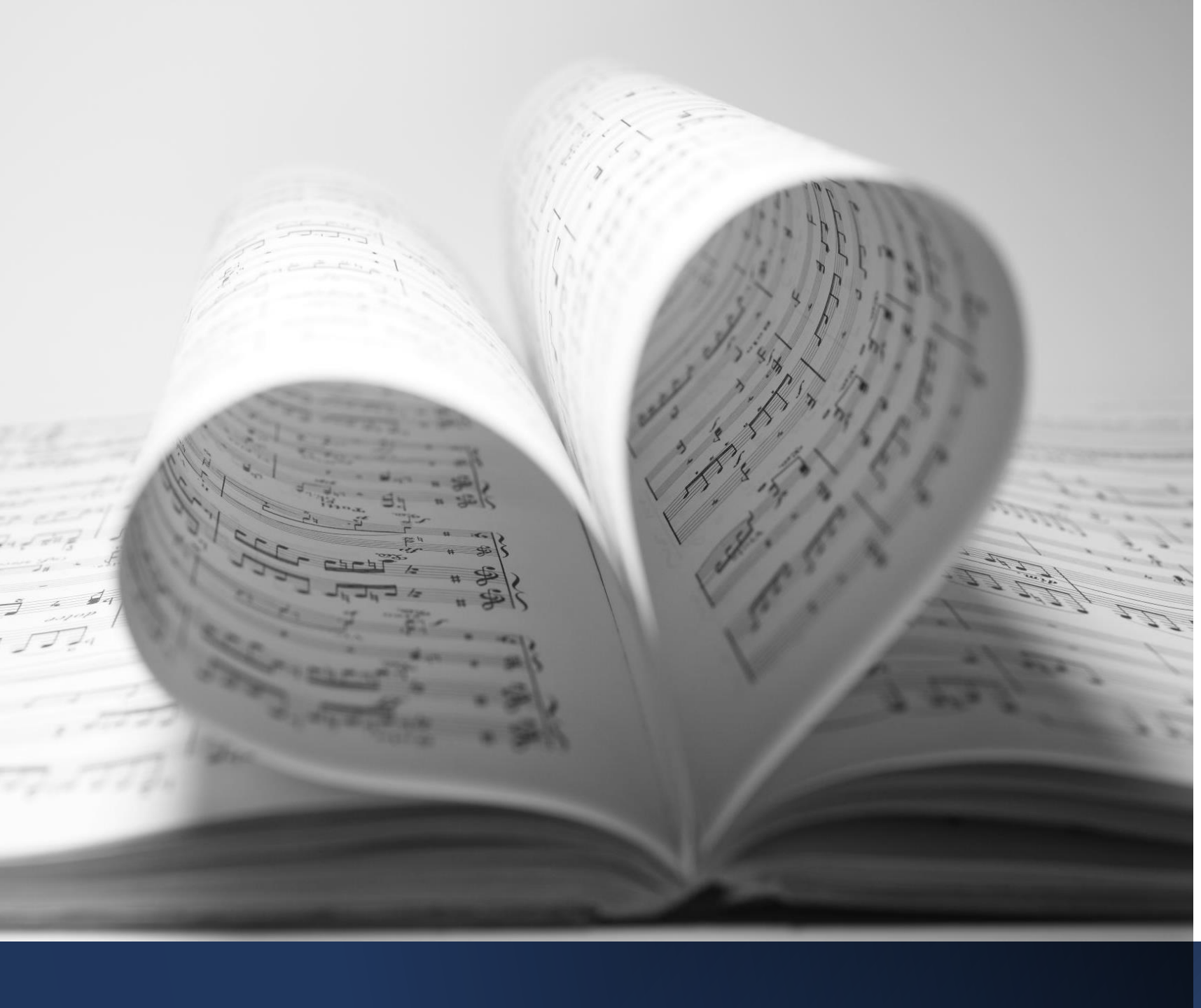
PHIL 351

Aaron Henry

Recognizing intelligence when we see it (II)



Readings

Get started on:

Cain, Chapter 1 §§4 to end

Dietrich et al. (pp. 45-55)

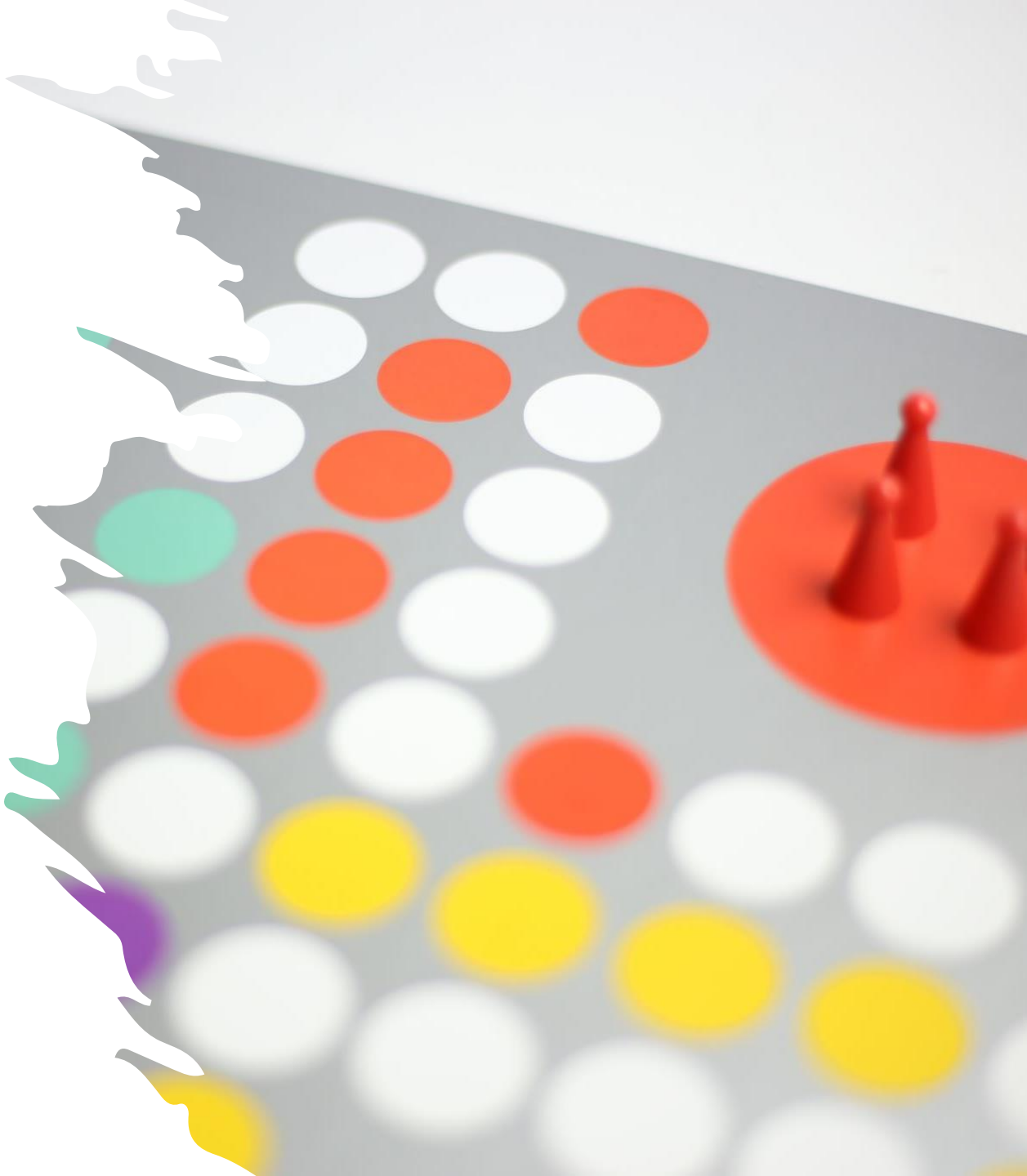
Cantwell-Smith, Chapter 1 ('Background')

Optional:

Kind, pp. 76-86

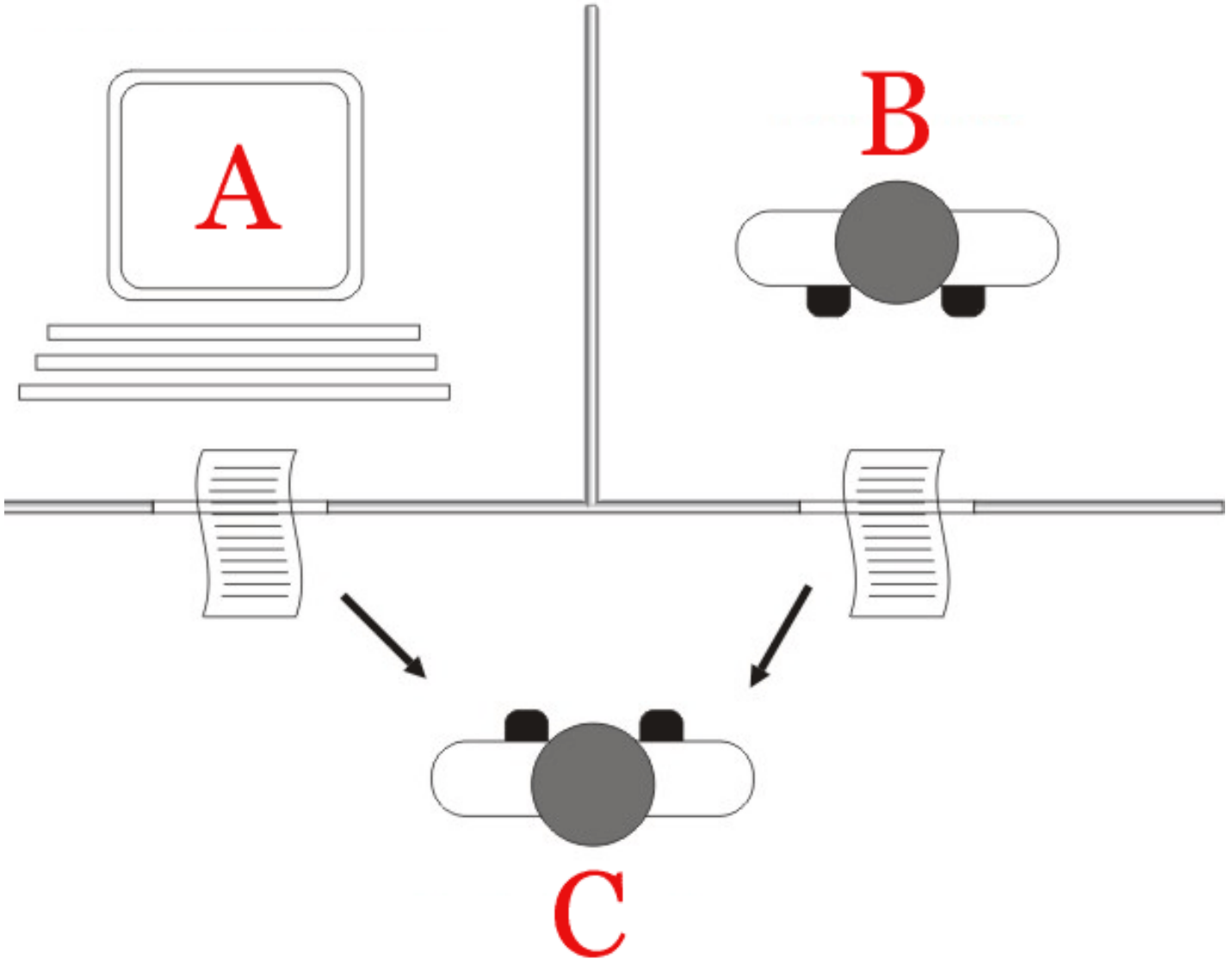
Plan

- Continue with Turing's imitation game (the Turing test)
- General characteristics of intelligent behaviour.
- Turing machines and the computational theory of mind.



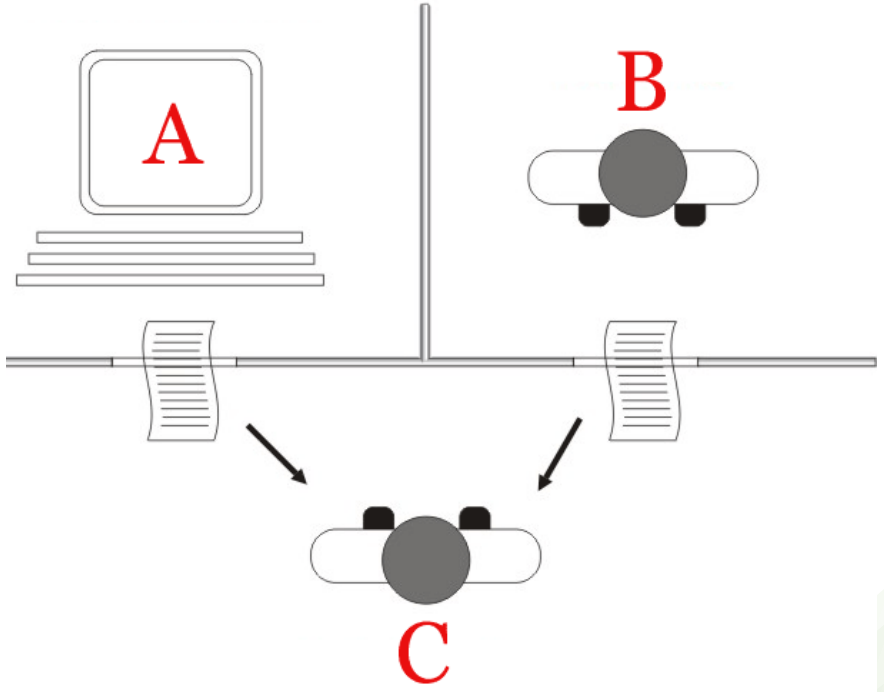
Assessing the Turing test as a test

- Many have wondered what Turing's exact intentions were in proposing the imitation game as a 'test' for intelligence.



Assessing the Turing test as a test

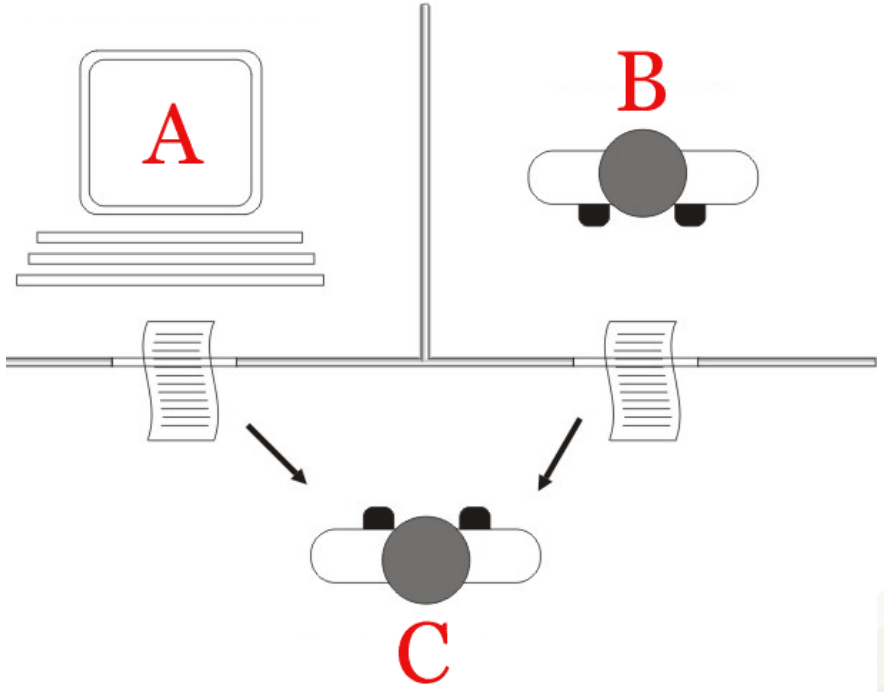
- Often read as proposing a kind of operational definition or (more plausibly) an empirically sufficient condition for intelligence.
- So interpreted, the test needs some filling in, since there are many ways of specifying its exact standards. For example, how will want to know:
  - How long are the conversations?
  - How often does the machine have to win?
  - How many judges must there be?



Assessing the Turing test as a test

We can distinguish different versions in response to different answers.

- **Minimum Turing Test:** Trick a single human being, at least once, into thinking you're human, after a brief and relaxed text conversation.
- **Maximum Turing Test:** Reliably perform at 70% at tricking multiple highly motivated and fairly clever judges into thinking you're human, no matter what searching questions they ask.

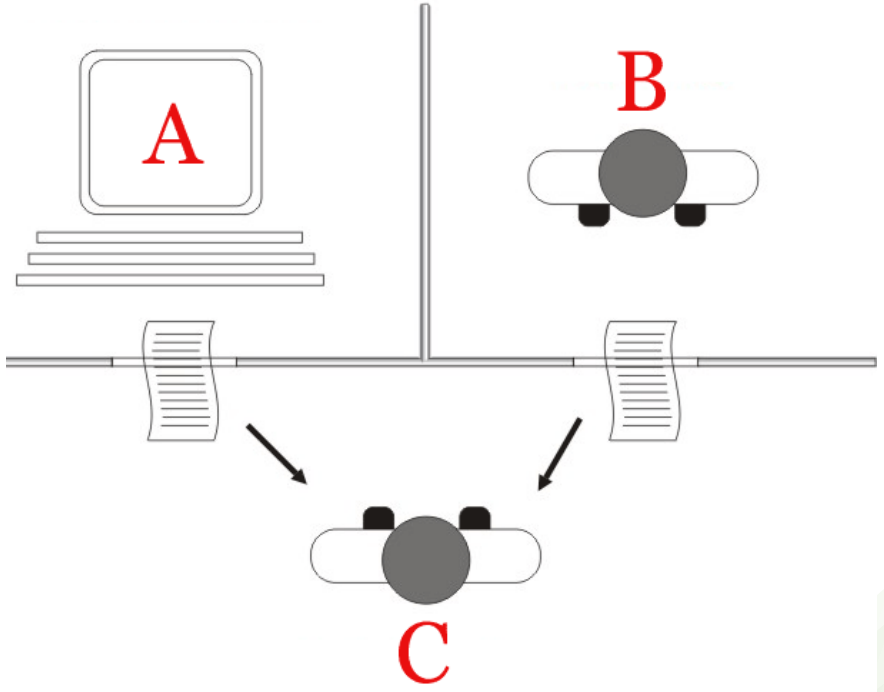


Assessing the Turing test as a test

Even after we distinguish ‘minimum’ and ‘maximum’ versions, questions remain. For example, can the judges be computer scientists?

Interestingly, Turing’s answer seems to have been ‘no,’ on the grounds that this would be unfair to the computer. Others say that the judges *should* be computer scientists.

- Cf. controversies around modern Turing tests like the Loebner Competition (Dietrich et al, p. 27 ff).



Assessing the Turing test as a test

Despite hiding the physical make-up of the competitors, could other sources of prejudice still seep in and bias the judge's verdict – e.g., subtle but noticeably different patterns of verbal behaviour?

Perhaps Turing intended to bypass this concern by having the task be an *imitation* task.

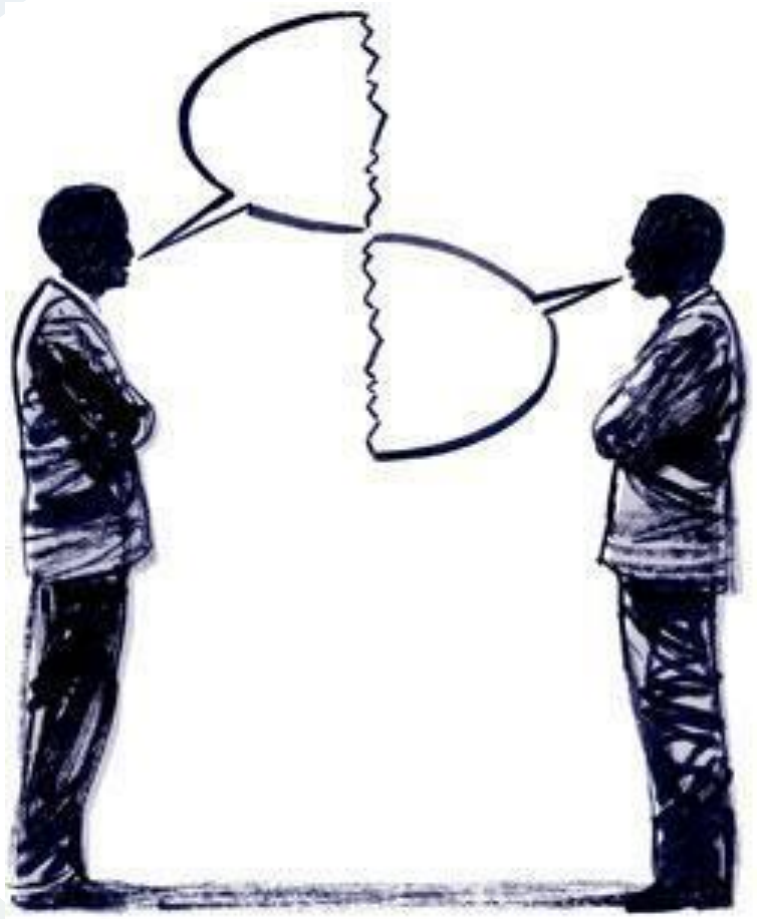
But, then, do we not hold the computer to a higher standard than the human that it competes against? Is that not prejudicial in much the same way it is prejudicial to judge intelligence by physical appearance? And isn't that precisely what we were trying to avoid?

Another interpretation of the Turing test

Dietrich et al argue that Turing never intended his 'test' be taken literally. His point, they suggest, is merely to entertain scenarios in which we are humbled by a computer's performance at a paradigmatically intelligent task.

More deeply, Turing is proposing a revision in how we use the word 'intelligence' (see §10). Specifically, Turing is proposing a constraint on an adequate definition of intelligence – namely, that it not be 'carbon chauvinist.'

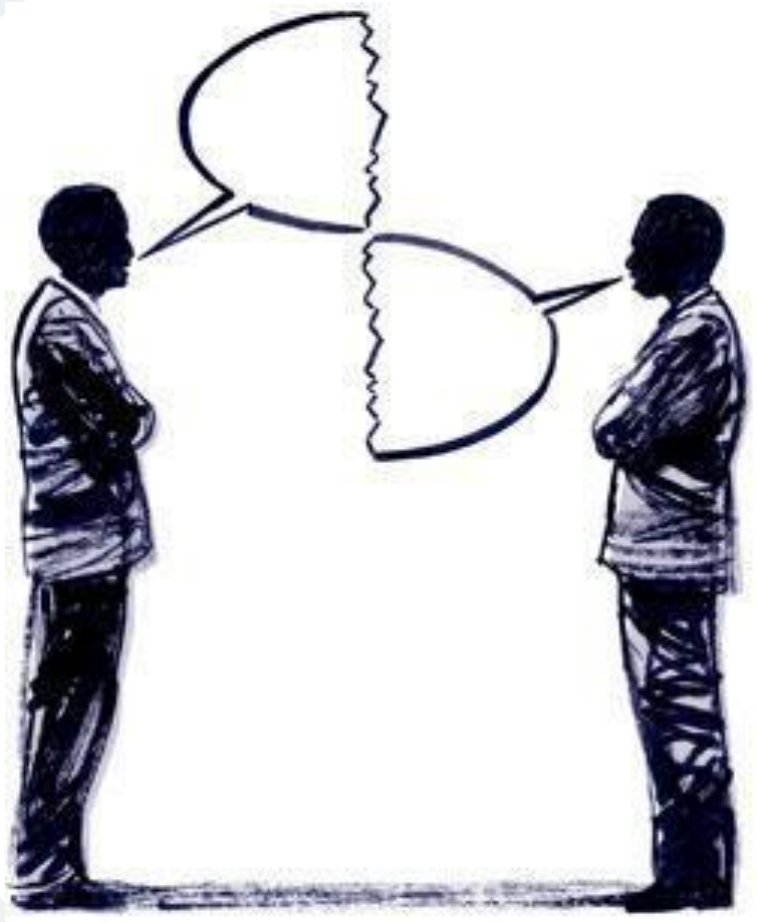
- Cf. 'meta-linguistic negotiation': what might appear as a purely 'verbal' (non-substantive) dispute at one level can actually be a substantive 'meta-linguistic' dispute about how a certain word *should* be used (cf. Plunkett 2015; Thomasson 2017).



So, no test then?

Okay, suppose we accept Turing's meta-linguistic proposal.

Are we to conclude that attempting to devise tests for the presence of intelligence is misguided? Should we give up?



A cue from William James

"Intelligence is a fixed goal with variable means of achieving it."

"The pursuance of future ends and the choice of means for their attainment are thus the mark and criterion of the presence of mentality" (Principles of Psychology, p. 21)"

General markers of intelligence

Using James' definition, we might fasten on certain canonical features of intelligent performances (cf. Fridland 2015):

- Goal-directedness
- Appropriateness
- Flexibility
- Transferability
- Manipulability

General markers of intelligence

Using James' definition, we might fasten on certain canonical features of intelligent performances (cf. Fridland 2015):

- Goal-directedness
- Appropriateness
- Flexibility
- Transferability
- Manipulability

A Sphexish tale

- The Sphex wasp has the (unfortunate) reputation of being a creature which *seems* intelligent but which is not *genuinely* intelligent.

A Sphexish tale

- What makes it *seem* intelligent: its behaviour is *appropriately goal-directed*.
- Why the appearance is (reputedly) misleading: the behaviour is highly *inflexible*.
- In fact, the Sphex may not be *quite* as inflexible as its reputation suggests (see Keizer 2013). But we'll work with the mythical Sphex.

Flexibility/response-freedom

- The (mythical) Sphex illustrates that appropriately goal-directed behaviour is not intelligent unless it is also *flexible*; i.e., unless responses are not mandated by the stimulus but, rather, selected from various possible responses that the creature could have performed ('response freedom').
- The converse is plausibly also true: flexible response, as such, is not intelligent unless it is also appropriately goal-directed – e.g., purely random behaviour is not intelligent.

Transferability

- If flexibility is the capacity to execute many different performances in a particular situation, transferability is the capacity to execute the same type of performance in many different situations, including novel ones.
- If flexibility concerns the ability to vary one's responses appropriately to the situation, transferability concerns the ability to redeploy one and the same competence in various different situations.
  - As before, what matters is not transfer per se, but *appropriate* transfer (e.g., sensitive in its implementation to relevant features of the new context).

Manipulability/endogenous control

- According to many theorists, paradigmatically intelligent task performances are endogenously controlled by the agent, rather than exogenously controlled by the stimulus.
  - 'top-down' vs. 'bottom-up'
  - active rather than passive

What makes verbal behaviour so intelligent?

- With these (likely non-exhaustive) criteria, we can arguably explain why verbal behaviour has struck so many as a gold standard of intelligence.
- Having a conversation is a goal-directed activity that is subject to various norms of appropriateness, not only strictly inferential relationships between the propositions expressed, but various pragmatic norms (e.g., Gricean 'maxims' of conversational implicature). The nature of the task places strong demands on flexibility, transferability, and manipulability. The result is often highly creative and novel.
    - Cf. Lady Lovelace's objection to Turing that a machine will never be able to *innovate* or create something *novel*.

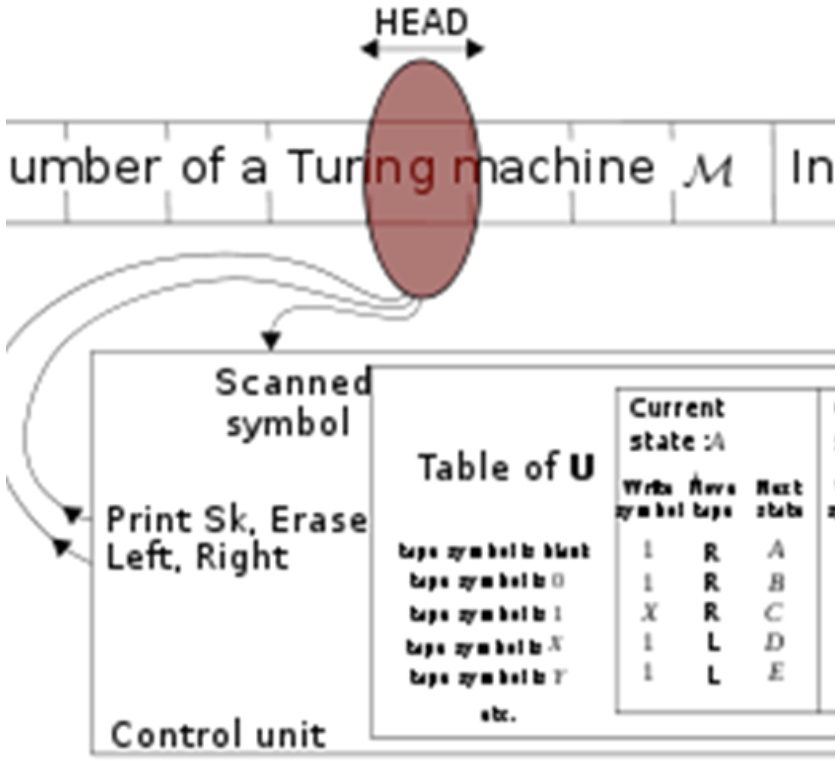
What makes verbal behaviour so intelligent?

- The same general criteria allow us to see human linguistic activity as just one among many possible behavioural expressions of intelligence. Specifically, they allow us to demarcate various kinds of intelligence and scales of intelligence.
- Showing versus telling: https://www.youtube.com/watch?v=cbSu2PXOTOc

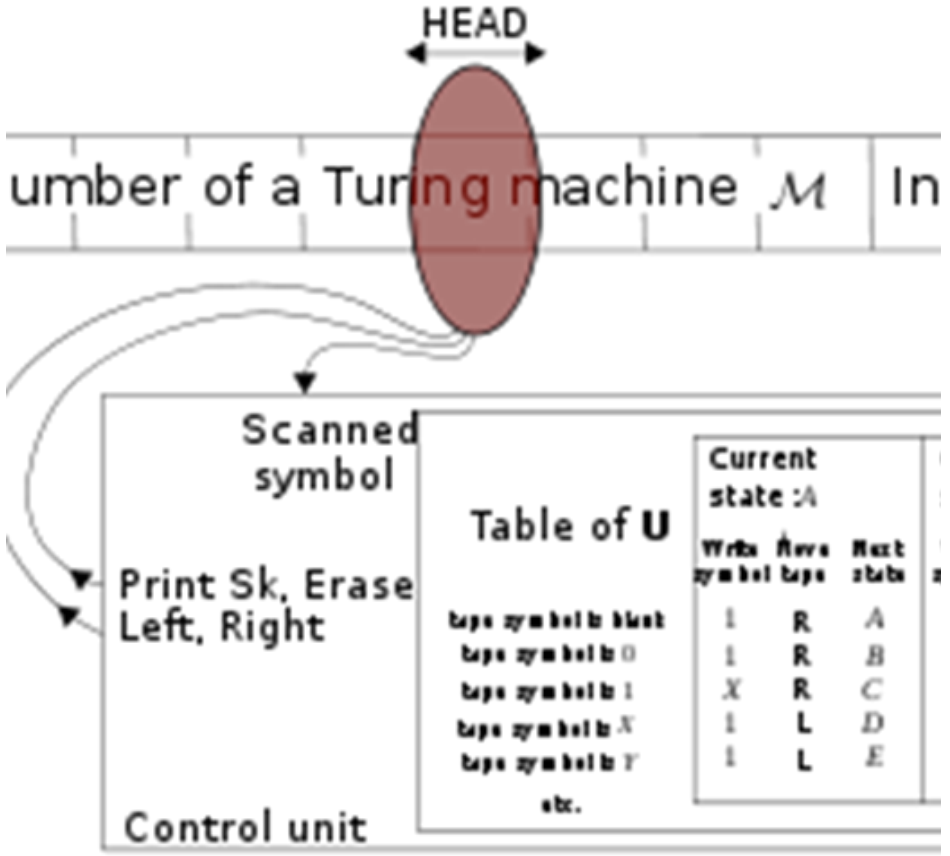
Turing machines and computation

‘Turing machines’

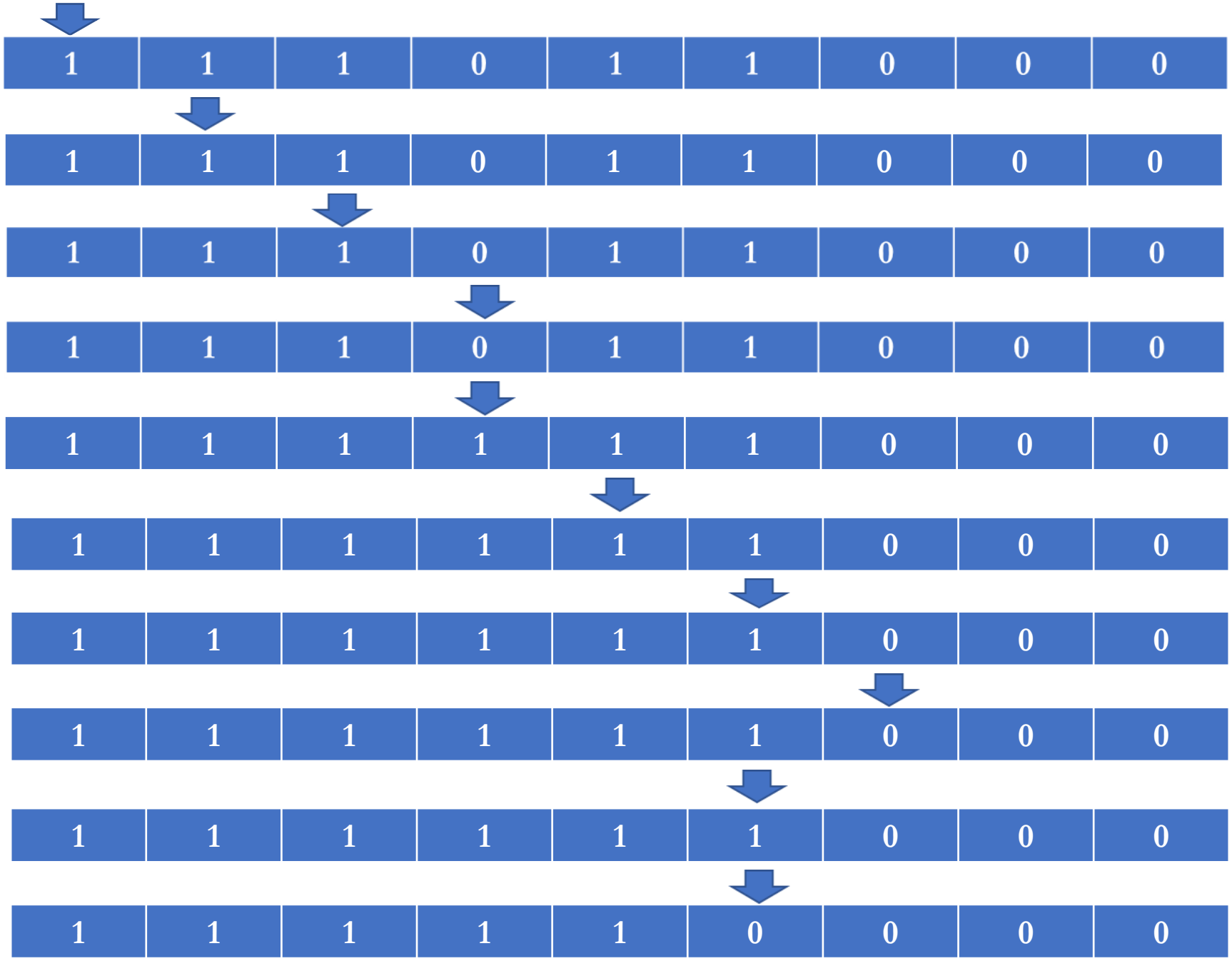
Turing introduced the notion of a Turing machine as a way of formalizing the notion of an ‘algorithm’ or ‘effectively computable’ function – intuitively, a recipe that can be followed mindlessly.



A ‘Turing machine’ is an ideal machine composed of an infinitely long piece of tape that is divided into cells. In each cell, a symbol can be written and erased. There is also a ‘machine head’ that, at any given time, is located over one of the cells on the tape and that can execute specific operations: it can *read* the symbol written on the cell, *erase* the symbol in the cell, *write* a new symbol in the cell, and *move* one cell to the left or right. Which operation it executes at a given moment is a function of its ‘machine table’ and of the machine’s internal state (of which there are finitely many). The machine table is the set of instructions that govern how the machine head behaves as a function of its internal state and the symbol inscribed in a particular cell.



Significantly, this leaves no room for ambiguity and requires no appeal to thought or judgment in the specification of the rules to be followed (exactly as required by the notion of an algorithm). It does this by being entirely *syntactically* governed.


1 1 1 0 1 1 0 0 0
1 1 1 0 1 1 0 0 0
1 1 1 0 1 1 0 0 0
1 1 1 0 1 1 0 0 0
1 1 1 1 1 1 0 0 0
1 1 1 1 1 1 0 0 0
1 1 1 1 1 1 0 0 0
1 1 1 1 1 1 0 0 0
1 1 1 1 1 1 0 0 0
1 1 1 1 1 0 0 0 0

Note what this simple machine just did: it successfully executed a *semantic* task: it calculated, in the language of binary, that the sum of 3 and 2 is 5. And yet, at no point did we require the machine have any understanding or insight into what the symbols *mean*. At each stage, its operations were physically sensitive to only purely syntactic properties of the symbols that it was manipulating (e.g., the shapes of the numerals '0' and '1').

Hence, a paradigmatically *mental* process (here, calculation) could be executed purely *mechanically*. And the same will be true for any function that meets certain very basic computability constraints.

‘Universal’ Turing Machines

Turing put these mathematical results to a number of further ends in ways that have been foundational for the field of computer science.

One such end was his proof of a ‘Universal Turing machine’, which can take the machine table of any specialized Turing machine (like the one we just described) and then run that machine. This is the core idea behind a general-purpose – i.e., programmable – digital computer with which we are all familiar today.