24.09 Minds and Machines
Fall 11 HASS-D CI

the Turing test

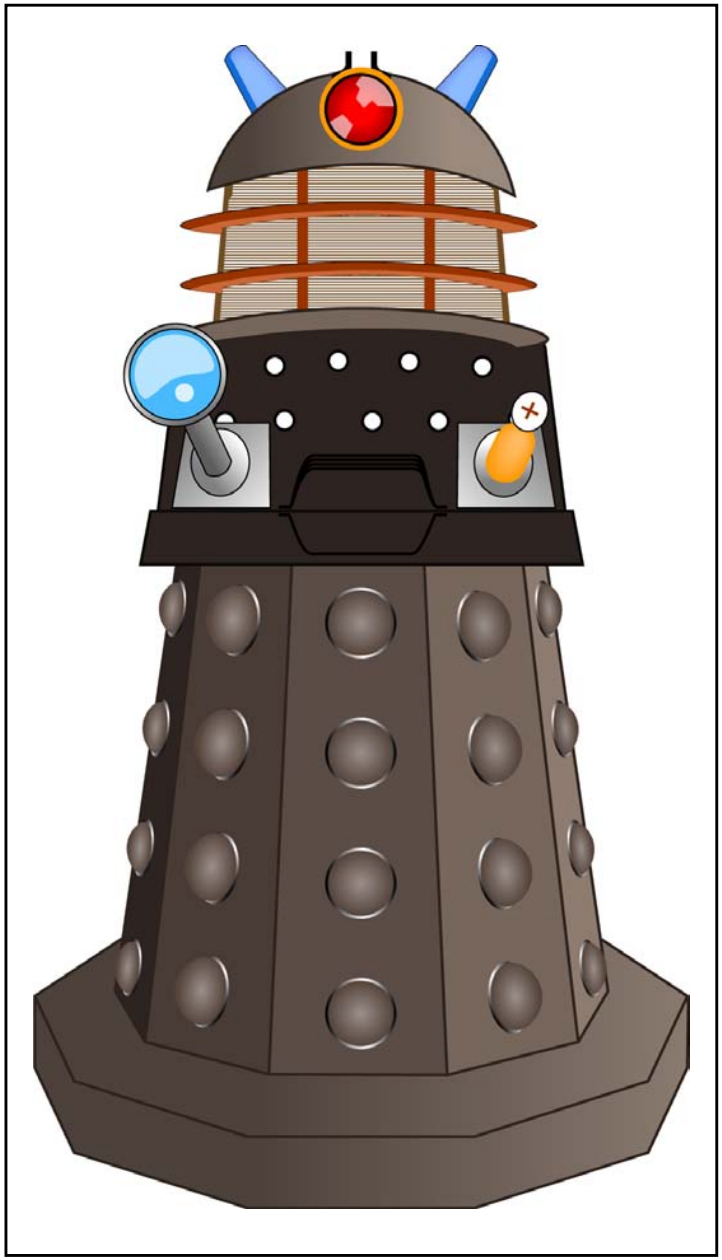


Image by MIT OpenCourseWare.

the robot reply

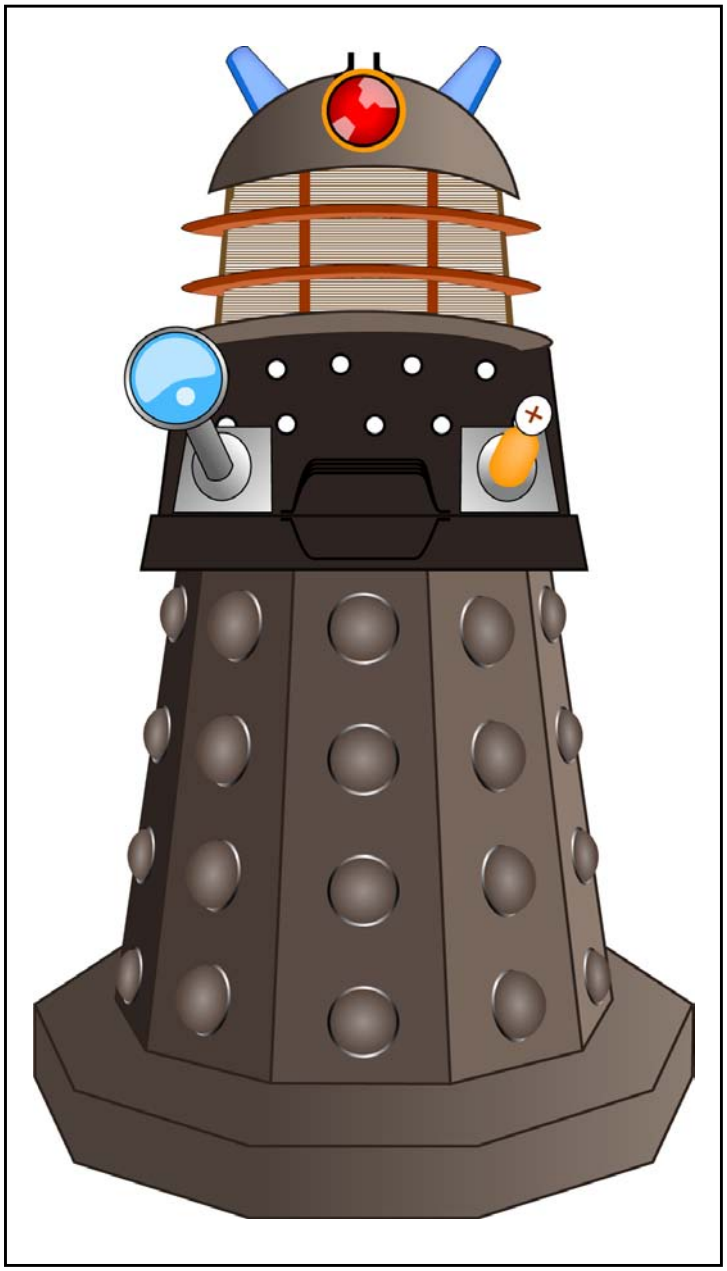


Image by MIT OpenCourseWare.

Inside a room in the robot's skull I shuffle symbols…As long as all I have is a formal computer program, I have no way of attaching any meaning to any of the symbols. And the fact that the robot is engaged in causal interaction with the outside world won't help me…

strong strong vs. weak strong AI

STRONG STRONG AI: there is a computer program (i.e. an algorithm for manipulating symbols) such that any (possible) computer running this program literally has cognitive states

WEAK STRONG AI: there is a computer program such that any (possible) computer running this program and embedded in the world in certain ways (e.g. certain causal connections hold between its internal states and states of its environment) literally has cognitive states

morals from the Chinese room

Searle's official argument against strong AI fails

but he does have a point, namely that merely implementing a program is arguably insufficient for (underived) intentionality

something else is needed—perhaps certain kinds of causal connections between the system and its environment

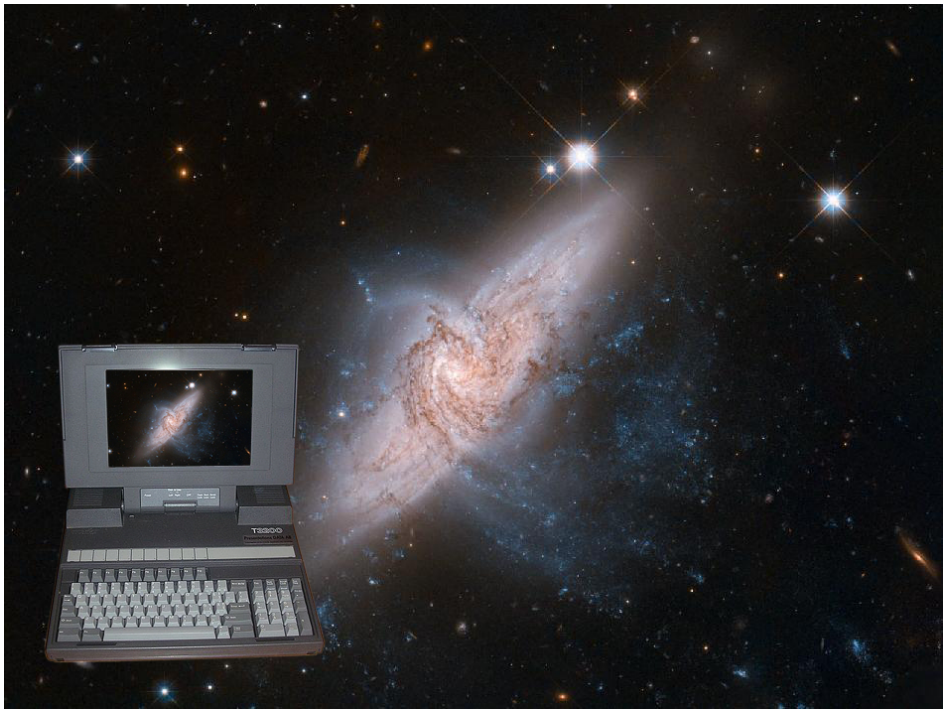


Image by MIT OpenCourseWare. Public domain photo courtesy of NASA.

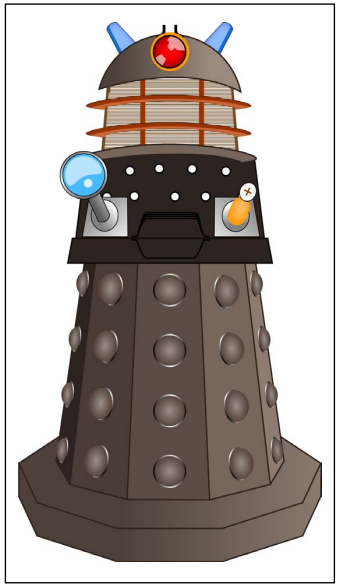


Image by MIT OpenCourseWare.

twin-Perry on twin-earth

running the alleged *intends-to-vote-for-Perry* program

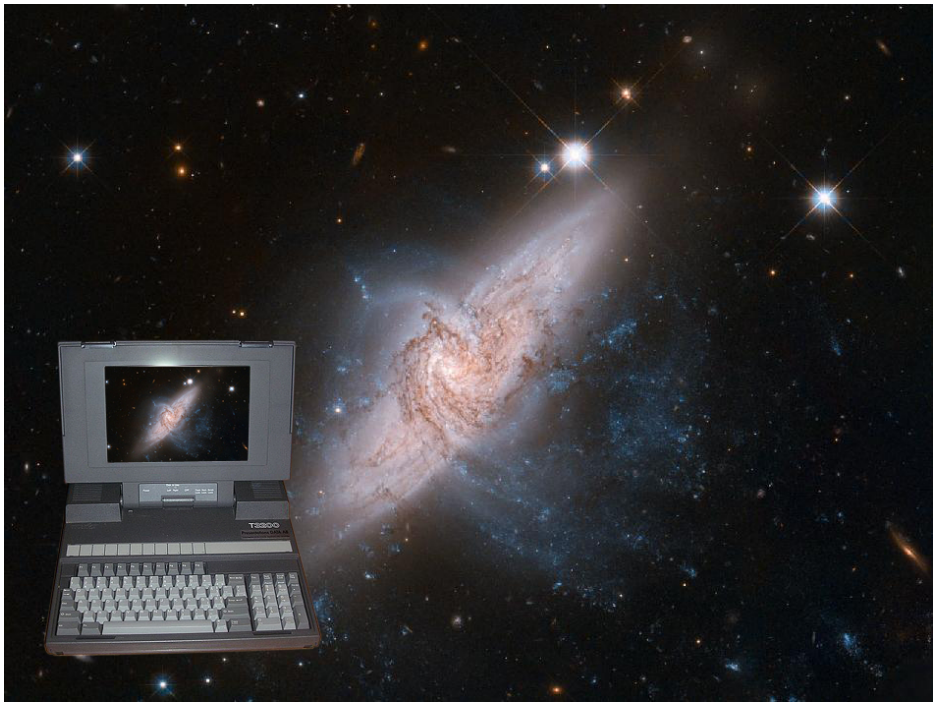
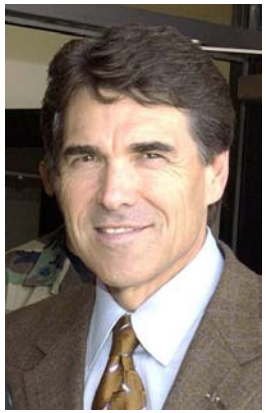
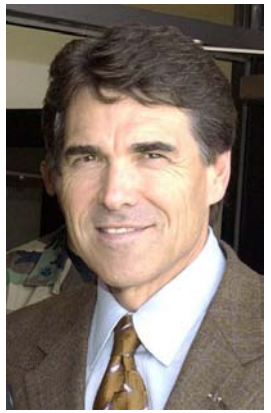


Image by MIT OpenCourseWare. Public domain photo courtesy of NASA.

why is the program about *Perry* rather than *twin-Perry*?



Perry



twin-Perry—lives on a planet in another galaxy

Searle makes a similar point

2. Syntax is not sufficient for semantics

3. Computer programs are entirely defined by their formal, or syntactical, structure

4. Minds have mental contents; specifically, they have semantic contents (from 'Can computers think?')

we'll return to this 'twin-earth' business later

‘Computing machinery and intelligence’ (1950)

I propose to consider the question, ‘Can machines think?’ This should begin with definitions of the meaning of the terms ‘machine’ and ‘think.’ The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words ‘machine’ and ‘think’ are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, ‘Can machines think?’ is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

the imitation game: man vs. woman

Image by MIT OpenCourseWare.

X

Image by MIT OpenCourseWare.

Y

C: 'Will X please tell me the length of his or her hair?'

X (man A, imitating a woman): 'My hair is shingled, and the longest strands are about nine inches long.'

(1950 is a long time ago)

the imitation game: human vs. machine

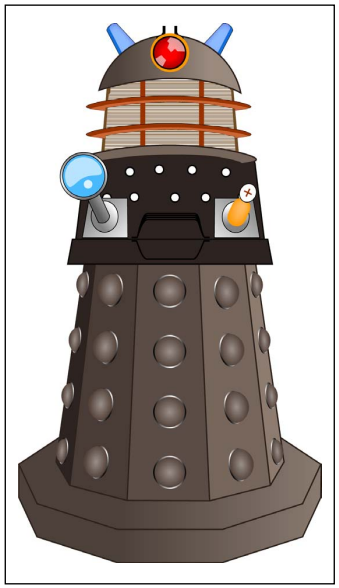


Image by MIT OpenCourseWare.

X

Image by MIT OpenCourseWare.

Y

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?'

but what is the replacement *for*?

this may not fit Turing's intentions, but according to Block the Turing test is intended to provide:

'...conceptual clarification. Turing was famous for having formulated a precise mathematical concept that he offered as a replacement for the vague idea of mechanical computability. The precise concept (computability by a Turing machine) did everything one would want a precise concept of mechanical computability to do. No doubt, Turing hoped that the Turing test conception of intelligence would yield everything one would want from a definition of intelligence without the vagueness of the ordinary concept.'

compare and contrast

f is computable if and only if *f* is Turing-computable

x is intelligent if and only if x can pass the Turing test

> the 'only if' direction of the latter is surely false, because it rules out intelligent agents who can't use language (e.g. chimps)

question:

is the proponent of Strong AI endorsing the claim that passing the Chinese-understanding Turing test is sufficient for understanding Chinese, etc?

x understands Chinese if x can pass the Chinese-understanding Turing test

answer: NO

another point: no such thing as 'the' Turing test

there are numerous Turing tests, corresponding to the numerous answers to these questions:

who are the judges?

how long is the test?

what's the subject matter?

the Loebner prize

Image removed due to copyright restrictions.

http://www.loebner.net/Prizef/loebner-prize.html

ELIZA

named after Eliza Doolitle in Shaw's *Pygmalion*

a 'Rogerian' chatbot therapist, developed at MIT in 1966

Image removed due to copyright restrictions.
A photograph of Joseph Weizenbaum (1923 - 2008).

talk to ELIZA

http://www.manifestation.com/neurotoys/eliza.php3

Image removed due to copyright restrictions.

a conflation

the claim that passing the test is sufficient for having a mind (etc.) should strike us as suspect

it seems to conflate:

(a) we have have excellent evidence for p

with

(b) p is true

we will return to this when we talk about 'behaviorism'

Block's Aunt Bubbles machine and the sufficiency claim

in any event, for any Turing test, no matter how demanding, there is a machine that can pass it that seems not to have a mind

this is shown by Block's 'Aunt Bubbles' example

Image removed due to copyright restrictions.

the Aunt Bubbles machine

'The machine works as follows. The judge goes first. Whatever the judge types in (typos and all) is one of $A_1 ... A_n$. The machine locates the particular A, say A_{2398}, and then spits back B_{2398}, a reply chosen by the programmers to be appropriate to A_{2398}. The judge types another message, and the machine again finds it in the list of Cs that sprout below B_{2398}, and then spits back the pre-recorded reply (which takes into account what was said in A_{2398} and B_{2398}). And so on. Though the machine can do as well in the one hour Turing Test as Aunt Bubbles, it has the intelligence of a juke-box. Every clever remark it produces was specifically thought of by the programmers as a response to the previous remark of the judge in the context of the previous conversation.'

inside Aunt Bubbles

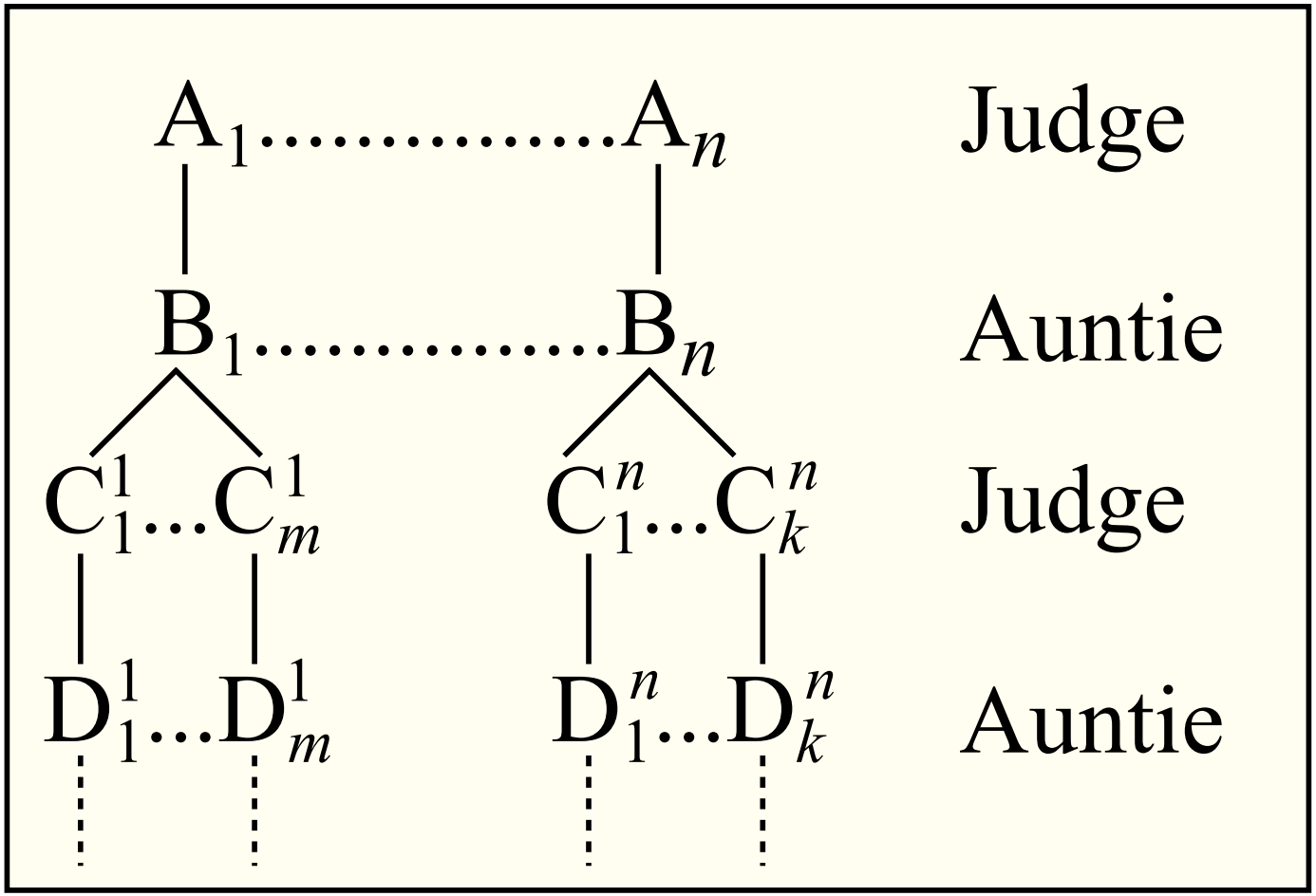


Image by MIT OpenCourseWare.

lesson from Aunt Bubbles (?)

minded creatures must have a certain sort of internal structure—being a giant lookup table isn't enough

Image removed due to copyright restrictions.

upcoming sessions

Scott Aaronson, CSAIL

dualism

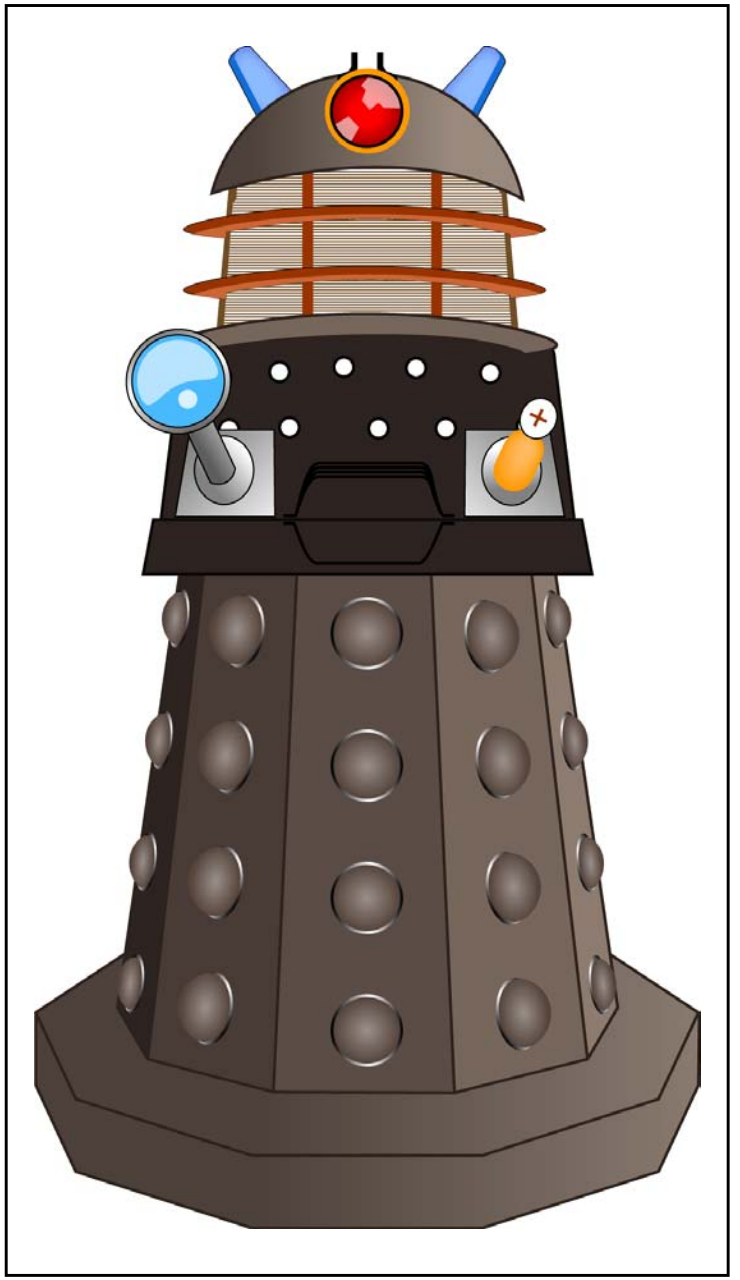


Image by MIT OpenCourseWare.

24.09 Minds and Machines

Fall 2011