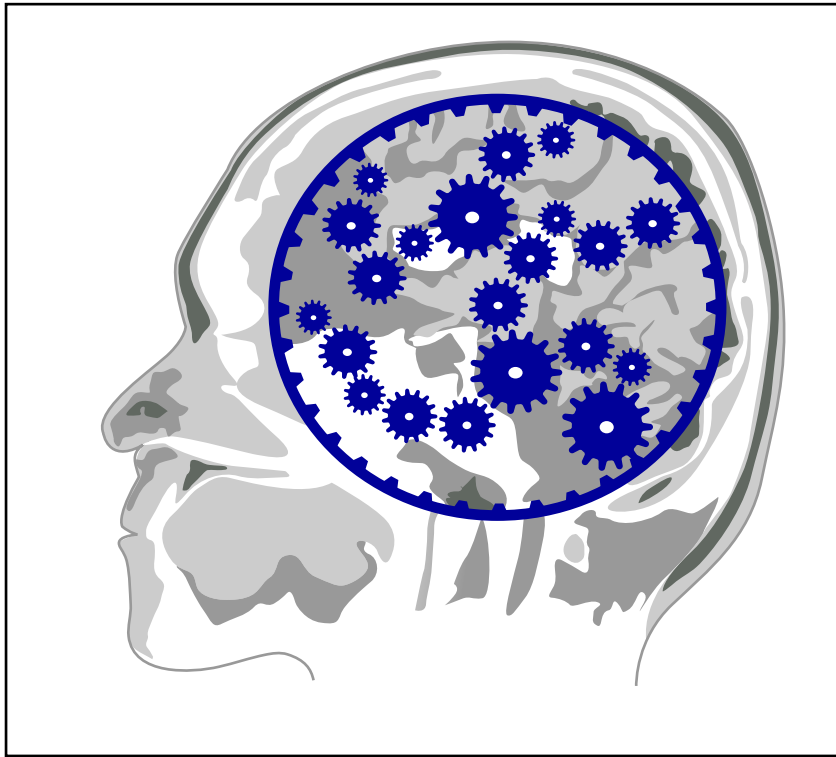# 24.09 Minds and Machines
spring 2007

Figure by MIT OCW.

- slides, announcements
  - see website
- recitations
- wed after class salon
- problem set, first argument analysis

intentionality

underived:

- the belief that Fido is a dog
- the desire for a walk
- the intention to use 'Fido' to refer to Fido

derived:

- the English sentence 'Fido is a dog'
- the Spanish sentence 'Fido es un perro'
-

PICK UP AFTER YOUR DOG

Figure by MIT OCW.

WEAK AI

- the principle value of the computer in the study of the mind is that it gives us a very powerful tool—e.g. it enables us to simulate various kinds of mental processes

Image removed due to copyright restrictions.

- cf. WEAK ARTIFICIAL METEOROLOGY

- WEAK AI is obviously correct (ditto WEAK AM)

STRONG AI

- an appropriately programmed computer literally has mental states (in particular, cognitive states)

- cf. STRONG AM—an appropriately programmed computer literally has meteorological states

- STRONG AI is disputable, and disputed by Searle

- STRONG AM, at least, is obviously false

the Chinese room

- a program: an algorithm (mechanical recipe) for transforming symbols into symbols

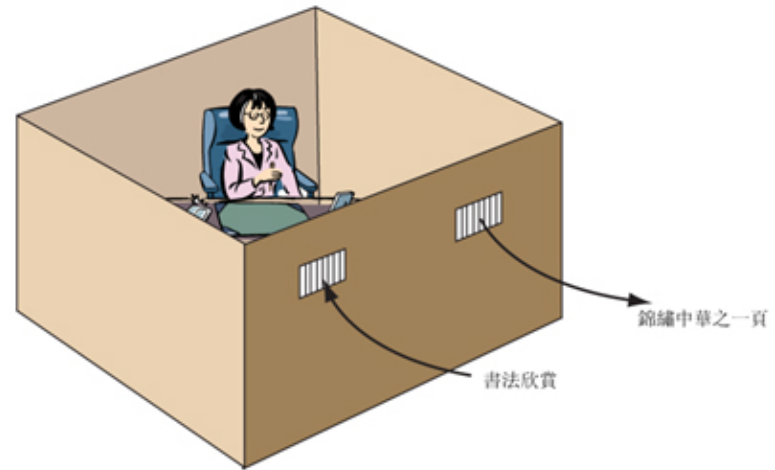- the thought experiment exploits the fact that computer programs can be "multiply realized"

Figure by MIT OCW.

digression on Turing machines

Image removed due to copyright restrictions.
Photograph of Alan Turing (1912-1954).

- wartime codebreaker, founder of computability theory

- invented "Turing machines"

- also invented the "Turing test" (more on this later)
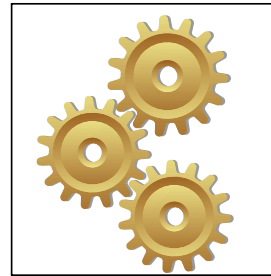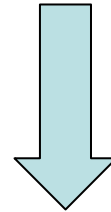
Turing machines

states: $S_1, S_2, \ldots, S_n$

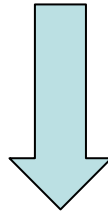Figure by MIT OCW.

head
{R, L, 1, 0, halt}

…10001111100001111101110000000
00…

tape

a simple Turing machine

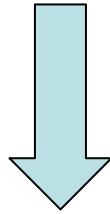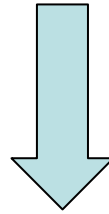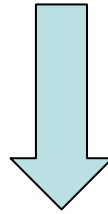| state / scanned cell	S_1	S_2
1	R, S_1	H
0	1, S_2	H

a simple Turing machine

0000000011111000000000000
00

S_1

a simple Turing machine

00000000111110000000000 00

S_1

a simple Turing machine

⬇

00000000111110000000000
00

S_1

a simple Turing machine

↓

0000000011111000000000000
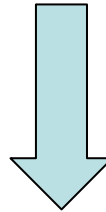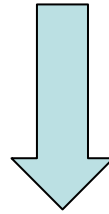00

S_1

a simple Turing machine

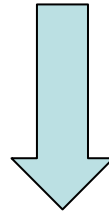

00000000111110000000000
00

S_1

a simple Turing machine
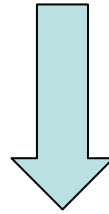
00000000111110000000000000

S$_1$

a simple Turing machine

00000000111111000000000
00

S_2

a simple Turing machine

00000000111111000000000
00

H

"…you behave exactly as if you understood Chinese, but all the same you don't understand a word of Chinese. But if going through the appropriate computer program for understanding Chinese is not enough to give <u>you</u> an understanding of Chinese, then it is not enough to give <u>any other digital computer</u> an understanding of Chinese"
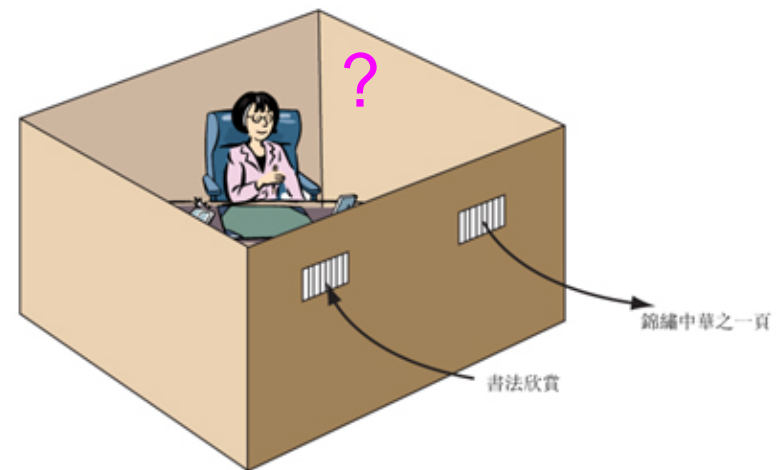
so, Strong AI is false

Figure by MIT OCW.

the systems reply

- the "whole system" understands Chinese, not Searle

- don't get hung up on "understanding Chinese"
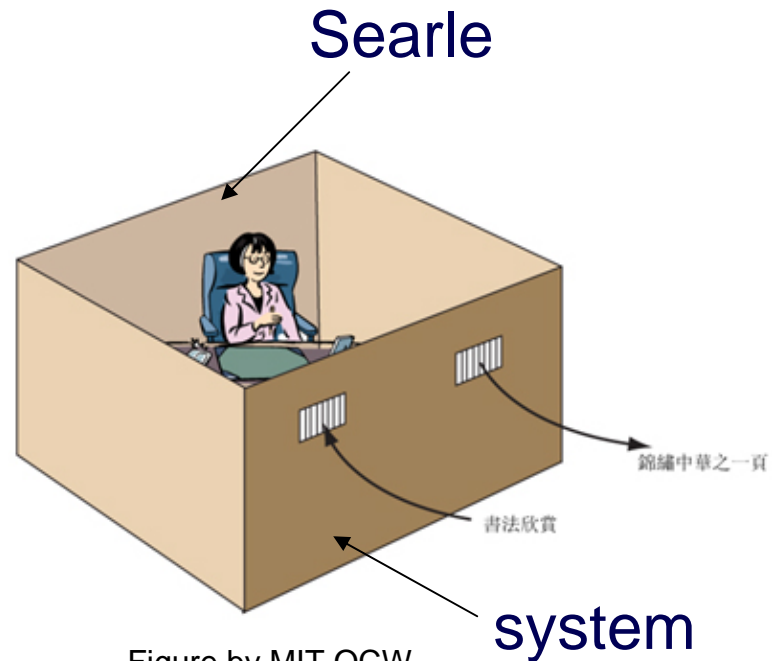 - it's simpler take the system to believe that it's raining in Beijing, e.g.

Searle

system

Figure by MIT OCW.

錦繡中華之一頁

書法欣賞

Searle's reply…

"…is quite simple: Let the individual internalize all of these elements of the system...he understands nothing of the Chinese, and a fortiori neither does the system, because there isn't anything in the system that isn't in him"

(from Searle, "Minds, Brains, and Programs")
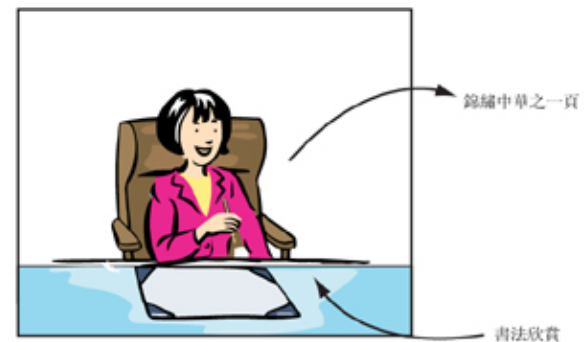
Searle (memorizes instructions)

Figure by MIT OCW.

Searle's reply…

… appears to rely on the
mistaken principle that if *x* is
part of *y*, and *y* isn't *F*, then *x*
isn't *F*. (My liver is part of me,
and I don't weigh 1 pound,
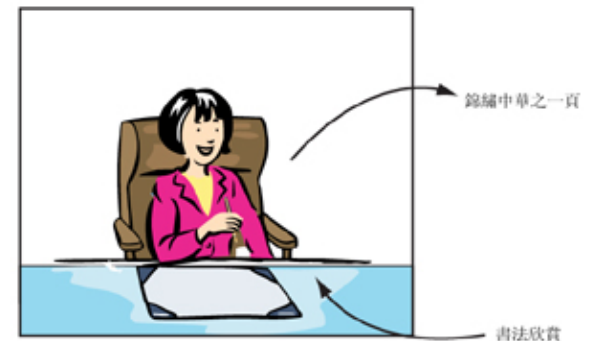but maybe my liver does.)

Figure by MIT OCW.

the robot reply

Image removed due to copyright restrictions.
Example robot.

"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 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

There is one aspect of Searle's case with which I am sympathetic. I have my doubts as to whether there is anything it is like to be the Chinese system, that is, whether the Chinese system is a <u>phenomenally conscious</u> system. My doubts arise from the idea that perhaps consciousness is more a matter of <u>implementation</u> of symbol processing than of symbol processing itself.
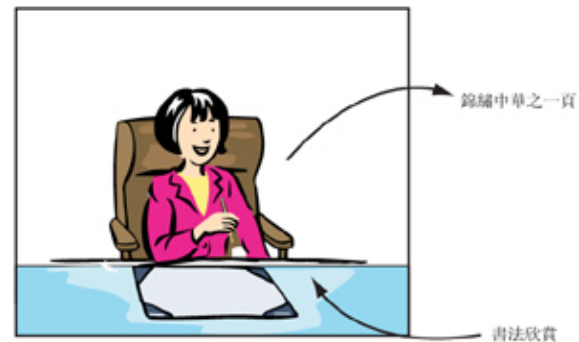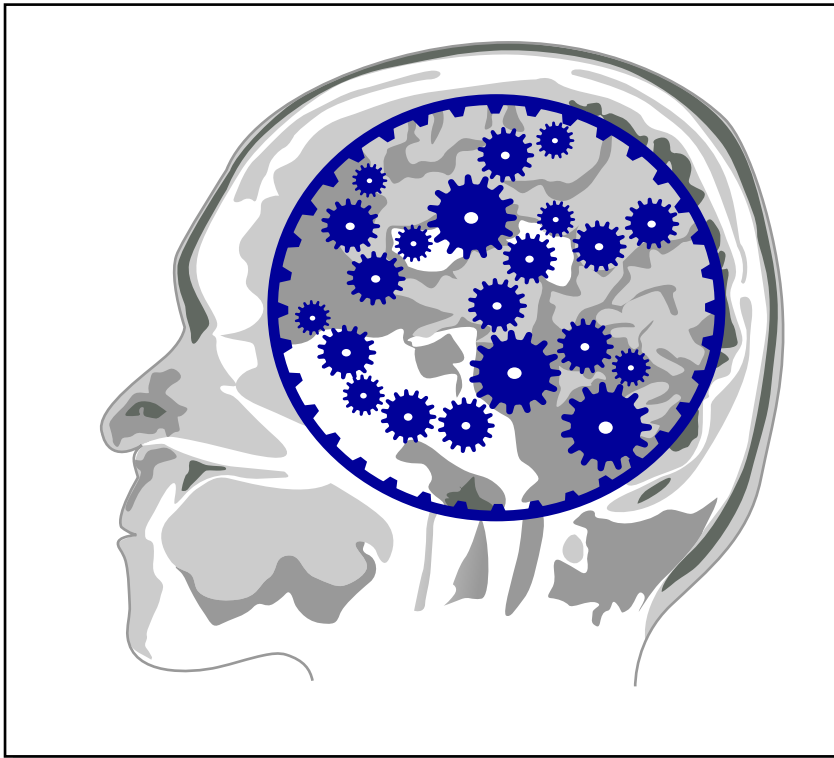
Block, "The mind as… "

Figure by MIT OCW.

24.09 Minds and Machines
spring 2007

Figure by MIT OCW.

- dualism

dualism

Image removed due to copyright restrictions.
Rene Descartes (1596-1650).

- *Meditations* (1642)

  "Wherein are demonstrated the existence of God and the Distinction of Soul from Body"

- *Principles of Philosophy* (1644)
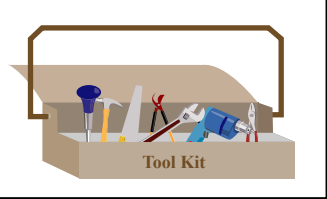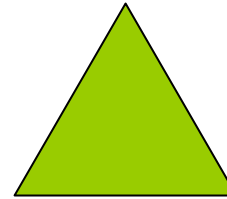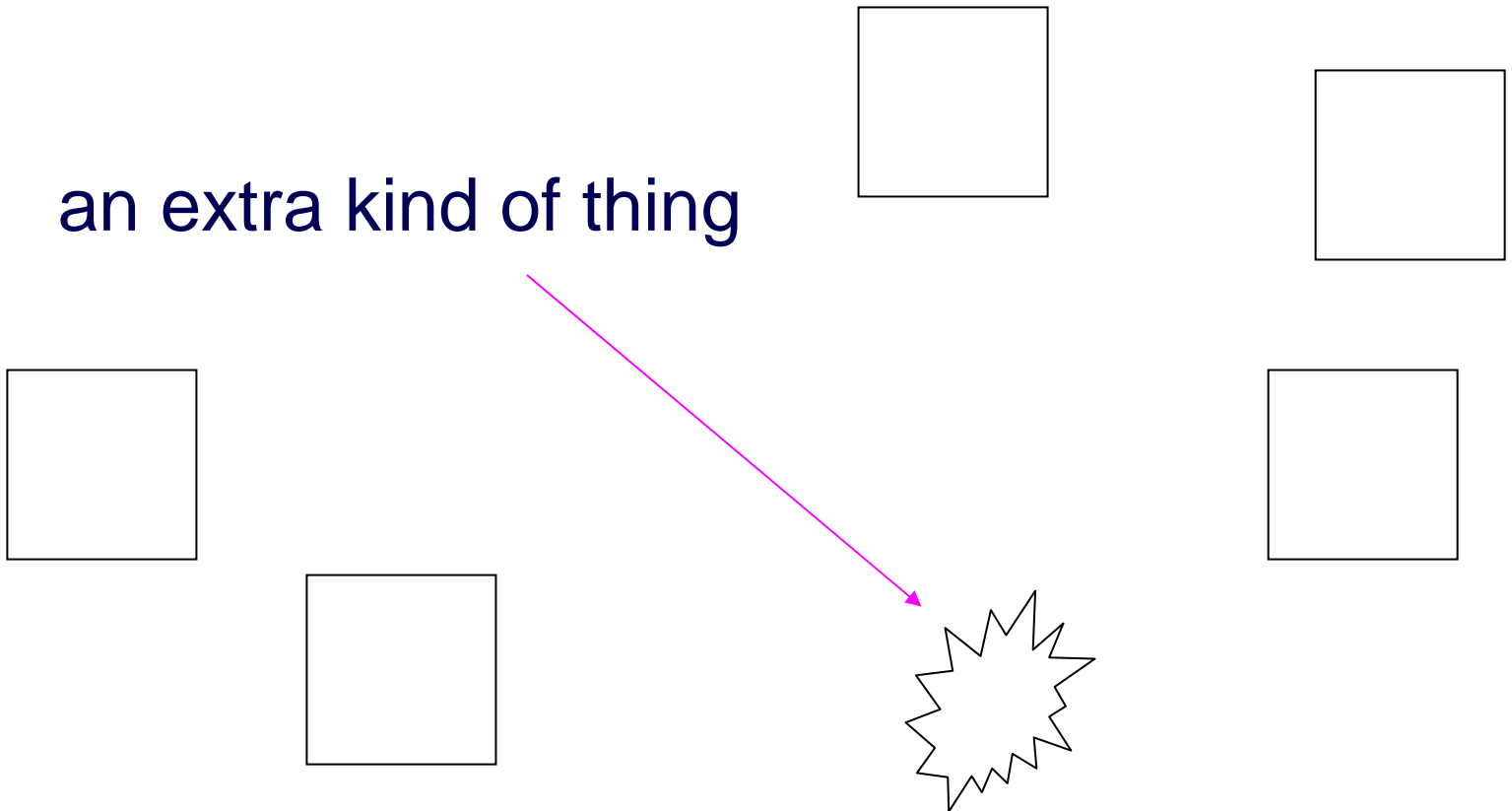
- *Passions of the Soul* (1649)

Figure by MIT OCW.

from the philosophical toolkit:
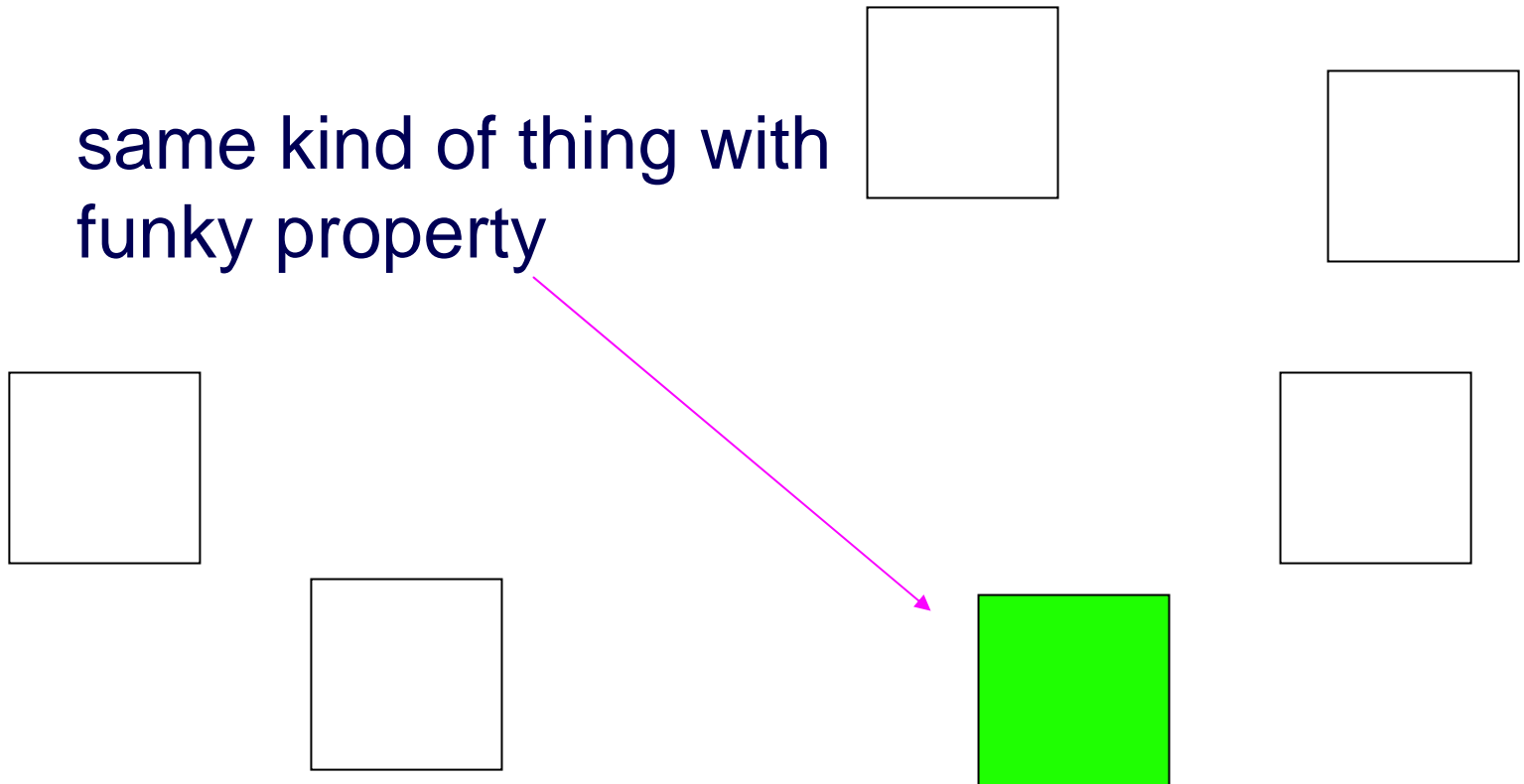
- <u>properties</u> and <u>particulars</u>

substance dualism

an extra kind of thing

property/attribute dualism
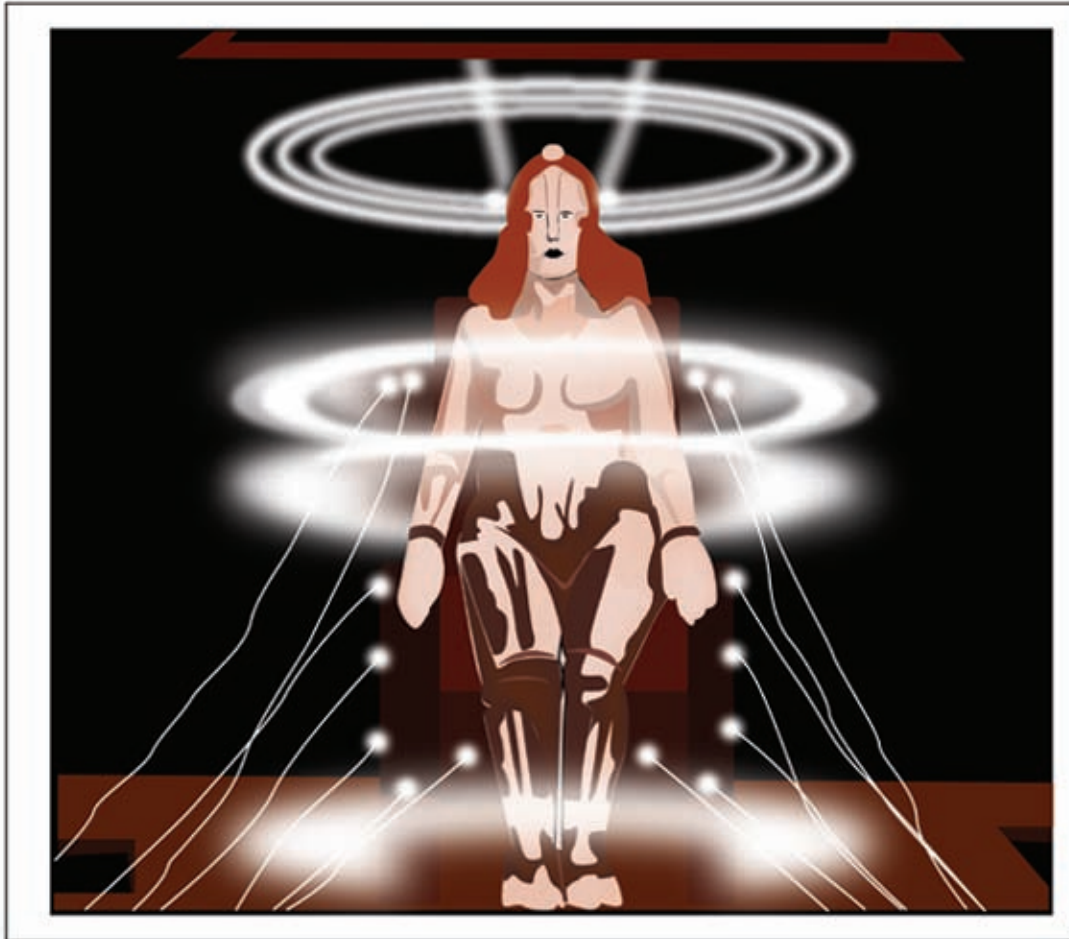
same kind of thing with
funky property

Minds and Machines
spring 2007

Figure by MIT OCW.

- next class on <u>tuesday</u>