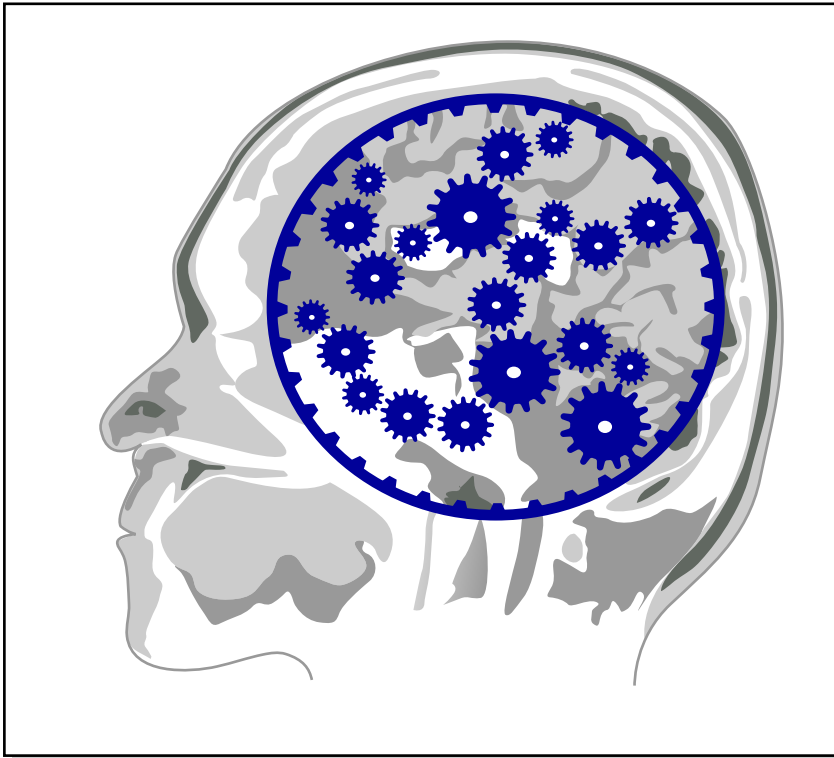


# 24.09 Minds and Machines

## spring 2007



- late problem sets
- syllabus adjustment
- externalism

Figure by MIT OCW.

recall our earlier discussion of the Chinese room:

## 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

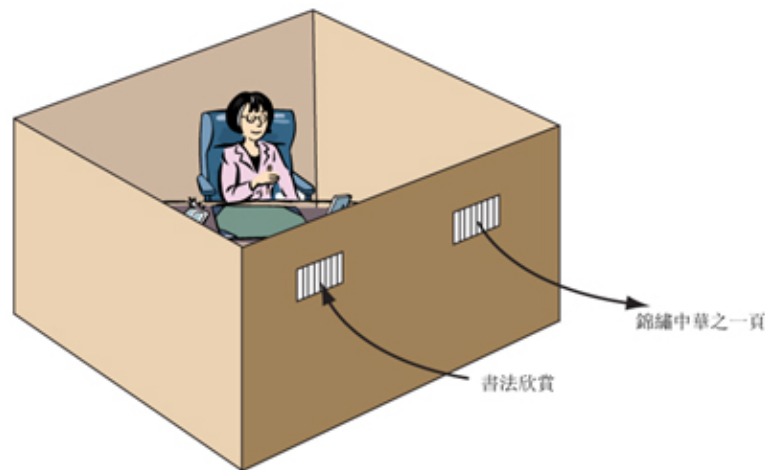


Figure by MIT OCW.

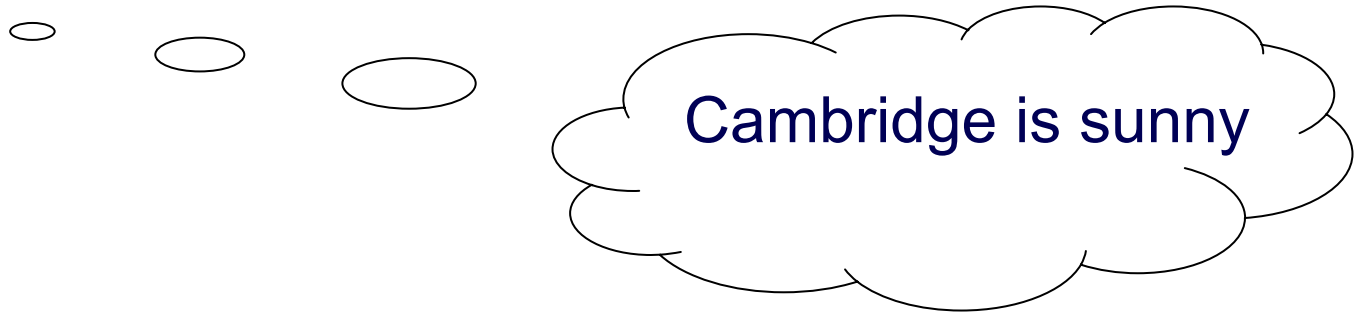
## WEAK STRONG AI

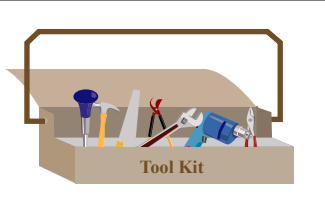
there is a computer program such that any (possible) computer running this program and embedded in the world in certain ways (e.g. certain causal connections hold between its internal states and states of its environment) literally has cognitive states

- if we were trying to develop a functionalist account of belief, would the relevant inputs and outputs involve distal objects in the system's environment (tomatoes, refrigerators, snow, etc.)?
- or would the inputs and outputs just be neural signals inside the creature's head?
- if the latter, then a system could believe that there are tomatoes in the fridge even if it's never had any tomato input

Image removed due to copyright restrictions.

Brain in a jar.





from the philosophical toolkit:

Figure by MIT OCW.

# extrinsic and intrinsic properties

perfect duplicates:

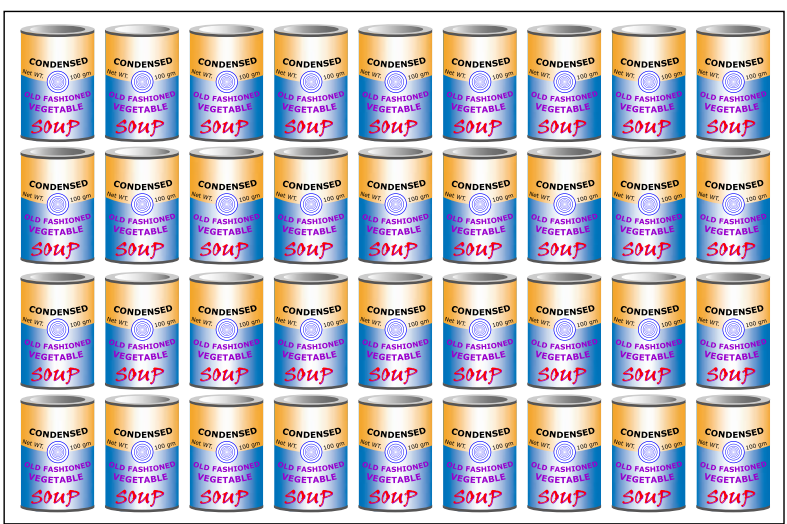


Figure by MIT OCW.

not-so-perfect duplicates:

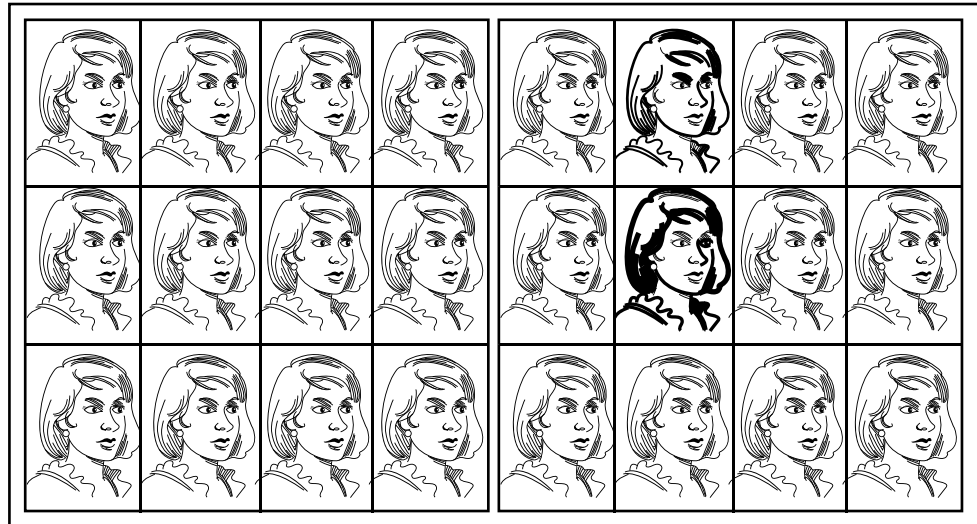
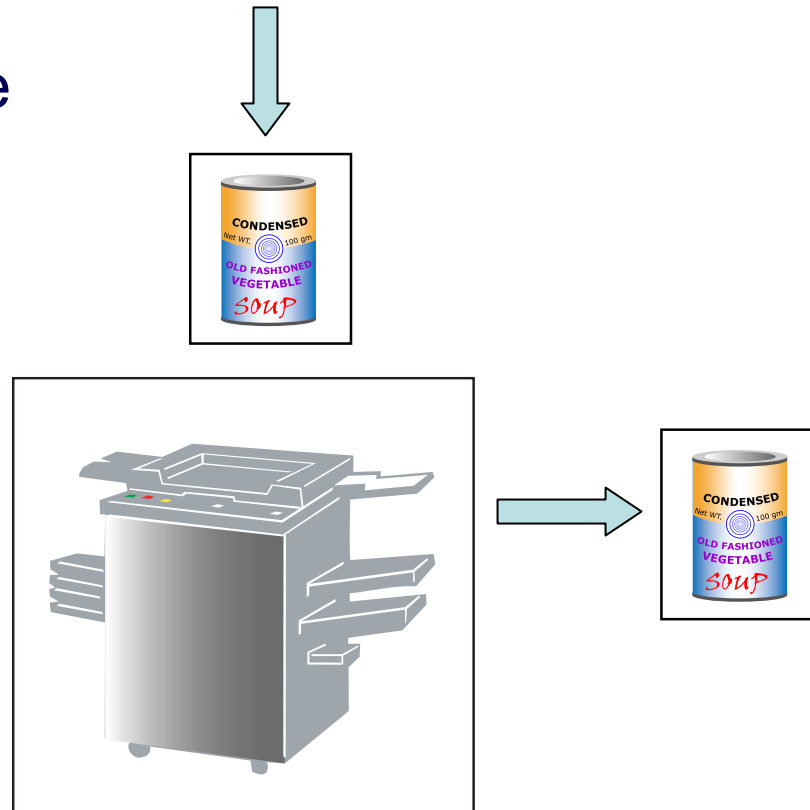


Figure by MIT OCW.

# extrinsic and intrinsic properties

the all-purpose  
duplicating machine



Figures by MIT OCW.

# extrinsic and intrinsic properties

- property  $P$  is intrinsic iff it is necessarily shared between perfect duplicates



Figures by MIT OCW.

- otherwise,  $P$  is extrinsic
  - there are many complications here, but we will ignore them; see the handout on properties and particulars

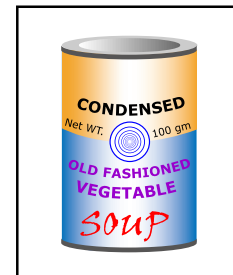
# extrinsic and intrinsic properties

- being cylindrical
- having mass 200g
- having red lettering (intrinsic)



Figures by MIT OCW.

- being a Campbell's soup can
- having weight 1960 N
- being in the cupboard (extrinsic)





# are mental properties intrinsic?

- yes, according to:
- Descartes (well, arguably—his dualism raises complications)
- the identity theory (taken as theory of all mental states, not just properties like being in pain)
- functionalism and behaviorism (on one natural way of spelling these theories out)
- commonsense (?)
  - however, contrast Lewis on pain (exercise: is being in human-pain intrinsic, on Lewis's account?)
- “thoughts are in the head!”

# are mental properties (of kind K) intrinsic?

- yes, according to internalism (about mental properties of kind K)
- no, according to externalism (about mental properties of kind K)
- we are about to look at some famous arguments for externalism about “propositional attitude” properties like wanting a glass of water, believing that Cambridge is pretty, etc.

# a “twin earth” thought experiment

Image removed due to copyright restrictions.

"Hubble Space Telescope Snaps Photo of Look-Alike Planet with Continents and Oceans -- Just Like Ours!" newspaper article.

# Cambridge and twin-Cambridge

Harvard Square

twin-Harvard Square



Hilary



twin-Hilary

Courtesy of Fredrick Jao. Used with permission.

# Cambridge and twin-Cambridge

“Cambridge is pretty”

“Cambridge is pretty”



Hilary



twin-Hilary

Courtesy of Fredrick Jao. Used with permission.

# different utterances

- Hilary's utterance is about Cambridge (not Twin-Cambridge, of which he has never heard)
- his utterance is true iff Cambridge is pretty
- the aesthetics of twin-Cambridge are totally irrelevant—if we imagine that twin-Cambridge is an imperfect duplicate of Cambridge (a twin Harvard Square, but exceptionally attractive elsewhere), then Hilary's utterance remains false, although twin-Hilary's utterance is true

# Cambridge and twin-Cambridge

“Cambridge is pretty”

“Cambridge is pretty”



Hilary



twin-Hilary

Courtesy of Fredrick Jao. Used with permission.

# different thoughts

- Hilary's thought is about Cambridge (not Twin-Cambridge, of which he has never heard)
- his thought is true iff Cambridge is pretty
- the aesthetics of twin-Cambridge are totally irrelevant—if we imagine that twin-Cambridge is an imperfect duplicate of Cambridge (a twin Harvard Square, but exceptionally attractive elsewhere), then Hilary's thought remains false, although twin-Hilary's thought is true



# “The meaning of ‘meaning’”

- this contains the original presentation of the “twin earth” thought experiment
- Putnam drew the conclusion that meanings aren’t “in the head” (i.e. aren’t intrinsic), not that thoughts aren’t in the head
- this paper is quite difficult, and we won’t discuss everything in it

# “The meaning of ‘meaning’”

- **intension** and **extension**
  - the extension of a term (e.g. ‘rabbit’, ‘creature with a kidney’) is “the set of things the term is true of”
  - so ‘creature with a kidney’ and ‘creature with a heart’ have the same extension
  - the intension of a term is its “meaning”, in the intuitive sense
  - so ‘creature with a kidney’ and ‘creature with a heart’ have different intensions

# “two unchallenged assumptions”

- 1) knowing the meaning of a term is just a matter of being in a certain psychological state
- 2) the intension of a term determines its extension (in the sense that sameness of intension entails sameness of extension)

# “two unchallenged assumptions”

- “I shall argue that these two assumptions are not jointly satisfied by any notion, let alone any notion of meaning. The traditional concept of meaning is a concept which rests on a false theory.”

# “methodological solipsism”

“no psychological state, properly so-called, presupposes the existence of any individual other than the subject to whom that state is ascribed.”

if this is right, then:

- being jealous of Smith
- seeing Jones
- knowing that the cat is on the mat
- remembering last night’s party

are not psychological states, properly so-called

- psychological states, according to **ms**, are “narrow”

the first “unchallenged assumption”,  
stated more precisely

- 1) knowing the meaning of a term  
(knowing that intension I is the  
meaning of term A) is just a matter of  
being in a certain narrow  
psychological state

# consequences of (1) and (2)

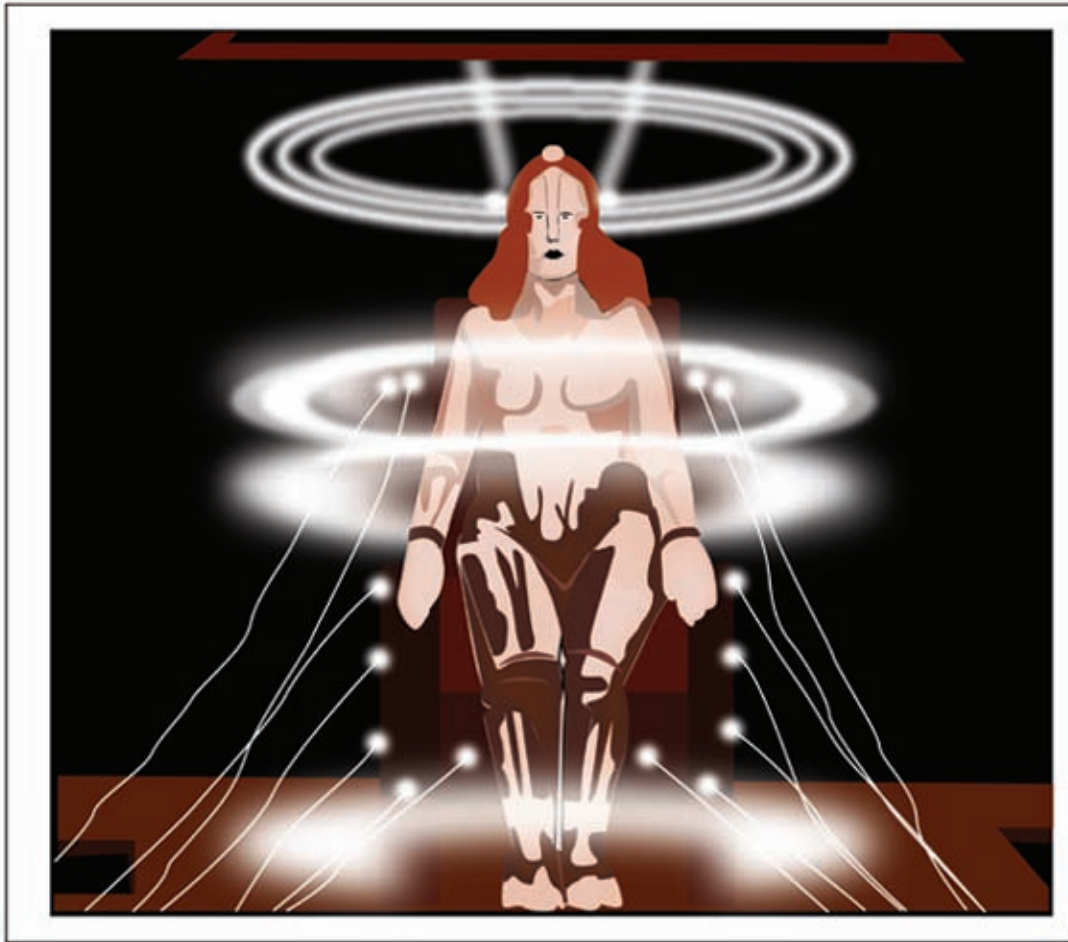
- suppose Oscar<sub>1</sub> and Oscar<sub>2</sub> are in the same narrow psychological states, and that Oscar<sub>1</sub> knows that intension I is the meaning of term A
- by (1), if Oscar<sub>1</sub> knows that I is the meaning of A, then Oscar<sub>2</sub> knows that I is the meaning of A
- if Oscar<sub>1</sub> and Oscar<sub>2</sub> both know that I is the meaning of A then Oscar<sub>1</sub> and Oscar<sub>2</sub> both use A with the same meaning or intension
- by (2), A as used by Oscar<sub>1</sub> has the same extension as A as used by Oscar<sub>2</sub>

- “It is this last consequence...that we claim to be false. We claim that it is possible for two speakers to be in exactly the same psychological state (in the narrow sense), even though the extension of the term A in the idiolect of the one is different from the extension of the term A in the idiolect of the other. Extension is not determined by psychological state.”



# Minds and Machines

spring 2007



- read Burge, Clark & Chalmers

Figure by MIT OCW.