

Minds and Machines

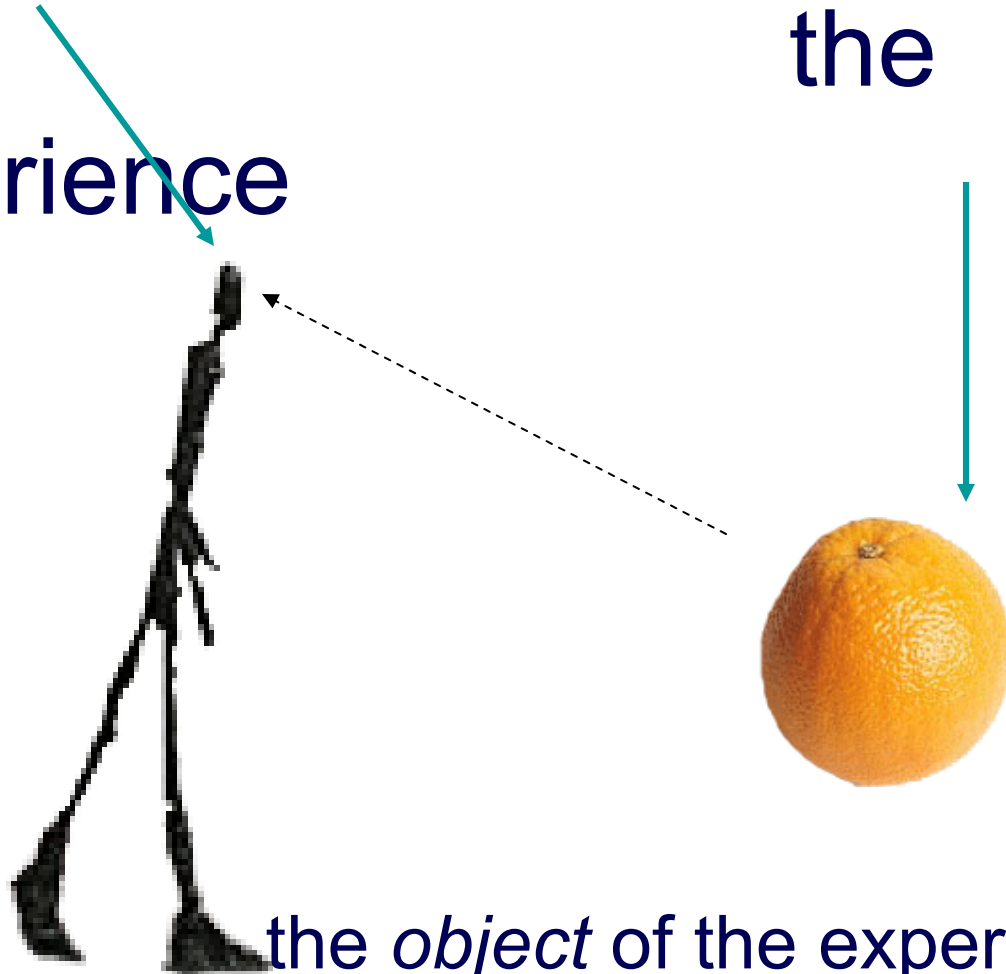
spring 2003

Functionalism

preliminaries

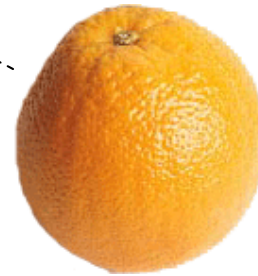
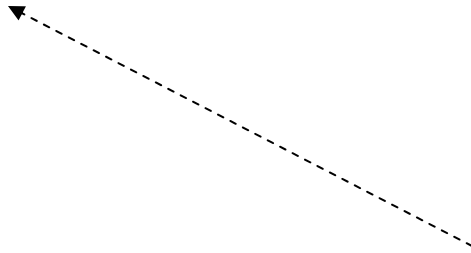
- the Chalmers collection
- Crane's book
- first paper topics in class #14

the experience vs. the object of
the
experience



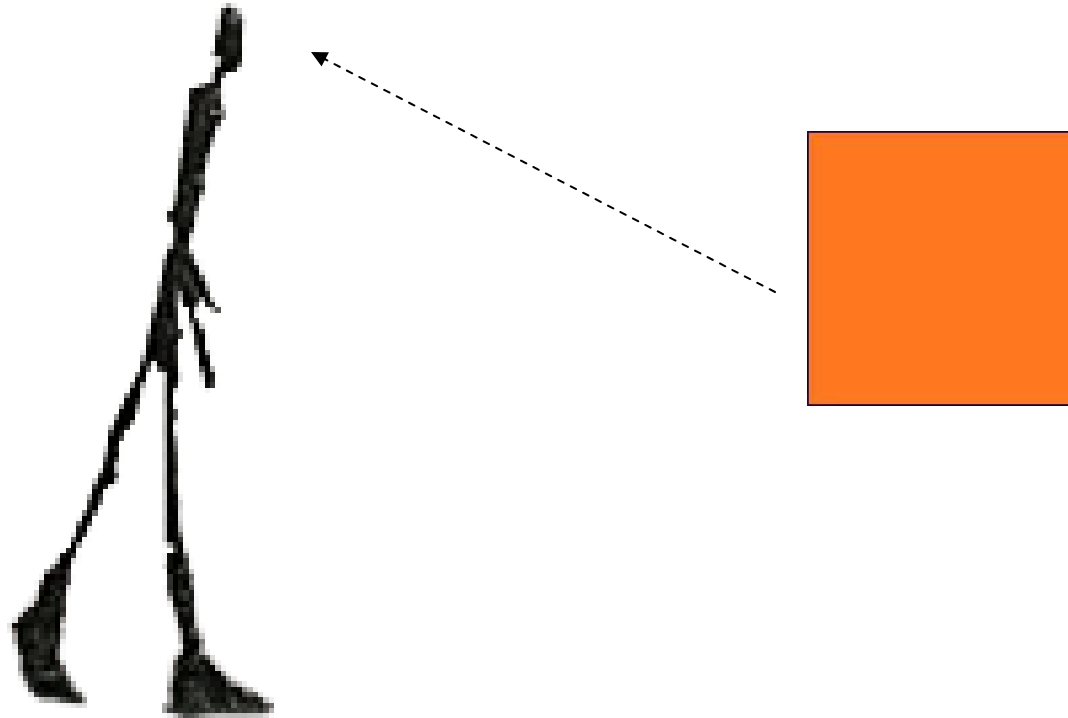
the *object* of the experience is colored orange, not the experience itself

the *experience* = such-and-such
brain process

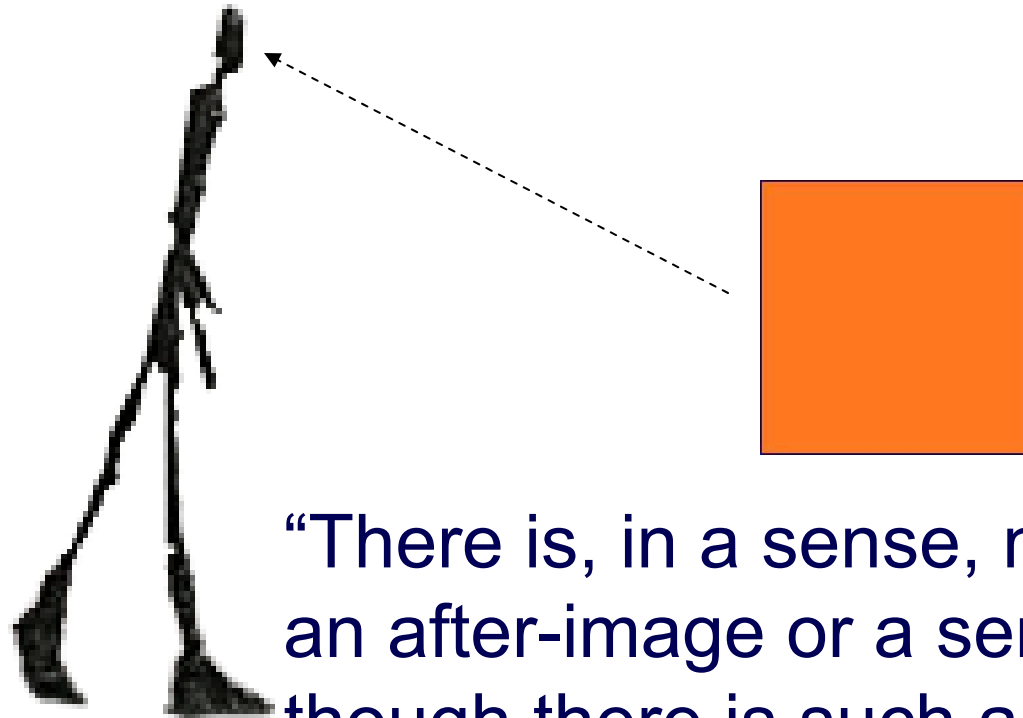


the identity theory does not identify
the *object* of the experience with a
brain process

Objection: the after-image is orange, but the brain process is not orange. So the after-image is not a brain process. (cf. “Objection 4”)



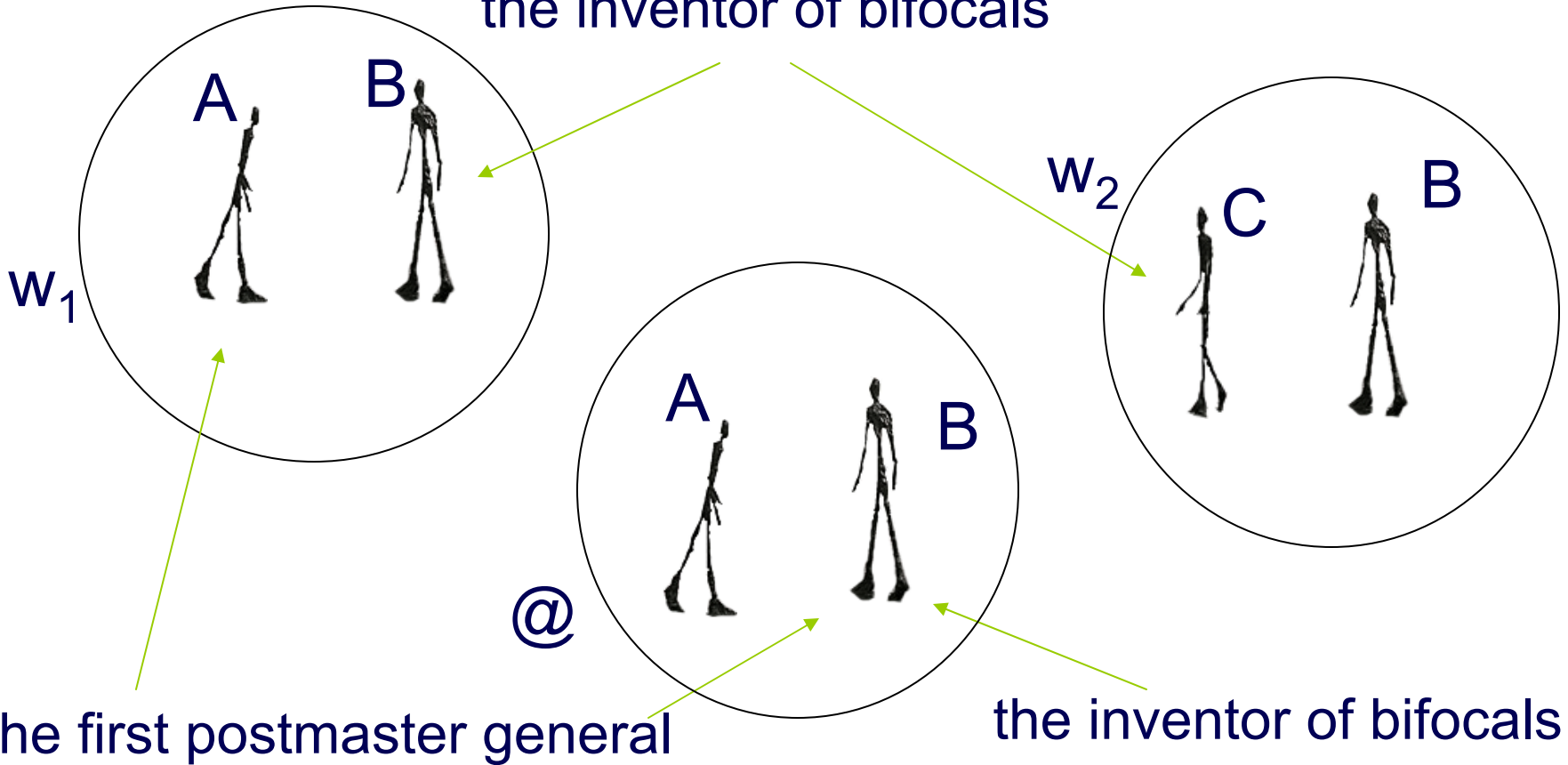
“I am not arguing that the after-image is a brain process, but that the experience of having an after-image is a brain process...”



“There is, in a sense, no such thing as an after-image or a sense-datum, though there is such a thing as the experience of having such an image”

some possible worlds

the inventor of bifocals



argument K_1

1. If I can clearly and distinctly conceive a proposition p to be true, then p is possible.
2. I can clearly and distinctly conceive that there is heat without mke (and vice versa). That is, I can clearly and distinctly conceive that the proposition that heat=mke is not true.

Therefore:

3. There is a possible world in which heat is not mke (it is not necessarily true that heat=mke).
4. If it's *true* that heat=mke, then it is *necessarily* true.

Therefore (from 3, 4):

5. Heat is not mke.

argument K_2

1. If I can clearly and distinctly conceive a proposition p to be true, then p is possible.
2. I can clearly and distinctly conceive that there is pain without c-fiber firing (and vice versa). That is, I can clearly and distinctly conceive that the proposition that pain=c-fiber firing is not true.

Therefore:

3. There is a possible world in which pain is not c-fiber firing (it is not necessarily true that pain=c-fiber firing).
4. If it's *true* that pain=c-fiber firing, then it is *necessarily* true.

Therefore (from 3, 4):

5. Pain is not c-fiber firing.

multiple realization

The [identity theorist] has to specify a physical-chemical state such that any organism (not just a mammal) is in pain if and only if (a) it possesses a brain of a suitable physical-chemical structure; and (b) its brain is in that physical-chemical state...it must be a state of the brain of any extra-terrestrial life that may be found that will be capable of feeling pain...

(Putnam, 77)

“replacement” scenarios

- a gradual replacement, one neuron at a time, with a chip that has the same input-output profile
- is this any worse than a cochlear implant?

lesson (?)

- minded creatures can in principle be made out of anything, provided certain structural constraints are met
- you couldn't make a mind out of Jello, but you could make one out of silicon chips

the Turing test

- see “[Computing machinery and intelligence](#)” (1950)
- is passing the test sufficient for having a mind/being intelligent/etc.?
- the question is underdescribed
- who are the judges?
- how long is the test?
- what’s the subject matter?

Alan Turing (1912-54)

the Turing test

- 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/could have excellent evidence for **p**
 - with
 - (b) **p** is true

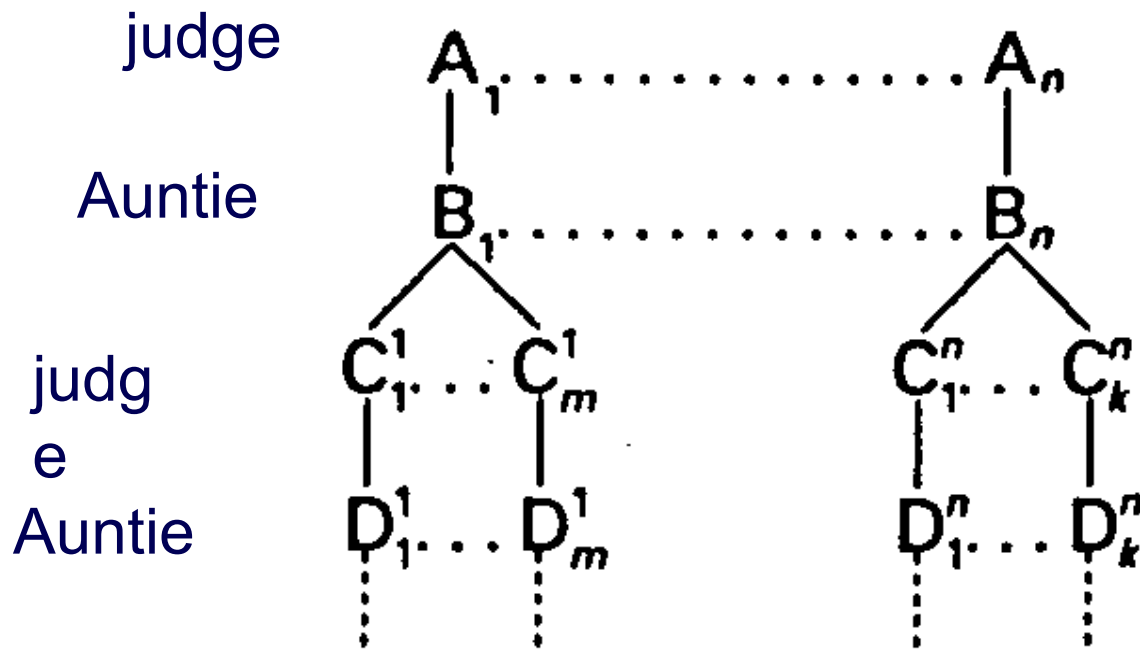
the Turing test

- 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

the Aunt Bubbles machine

The machine works as follows. The judge goes first. Whatever the judge types in (typos and all) is one of $A_1 \dots A_n$. 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 C s 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 the Aunt Bubbles machine



lesson (?)

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

lessons (?) from our discussion of behaviorism

- mental states are the inner causes of behavior
- the correspondence between mental states and behavior is many-many
 - (a) mental state M may produce different sorts of behavior, depending on the creature's other mental states
 - (b) the same behavior may be produced by different mental states

the mousetrap

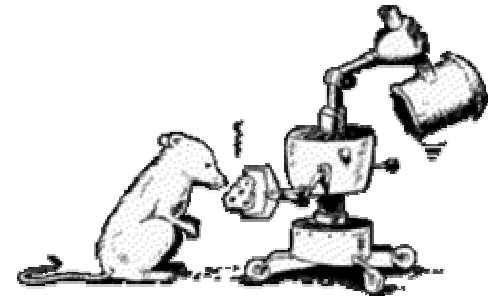
mousey
input



dead/trapped
mouse output



mousetraps can be “multiply realized”



the mousetrap

- the simplest sort of *functional kind*
- since no constraints are placed on its inner organization, it is also a *behavioral kind*






the 3-Coke vending machine

initial
state
s

output
for
25¢ input

next
states

$M_3 I_0$	$M_2 I_0$	$M_1 I_0$	$M_3 I_1$	$M_2 I_1$	$M_1 I_1$
"25¢"	"25¢"	"25¢"			
$M_3 I_1$	$M_2 I_1$	$M_1 I_1$	$M_2 I_0$	$M_1 I_0$	shut down

the 3-Coke vending machine

- it can be multiply realized
- a functional but not behavioral kind
- what *are* the states M_3 , I_2 , etc?
 - the table tells the whole story

Minds and Machines

spring 2003

- read Block, Lewis