Minds and Machines

spring 2003

Content:
intentionality and externalism, contd.

preliminaries

- adjustments have been made to the schedule

are mental properties intrinsic?

yes, according to:

- Descartes (well, arguably)
- 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(?)

“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 have looked at “twin earth” arguments for externalism about “propositional attitude” properties like wanting a glass of water, believing that Cambridge is pretty, etc.

Cambridge and twin-Cambridge

Hilary

twin-Hilary

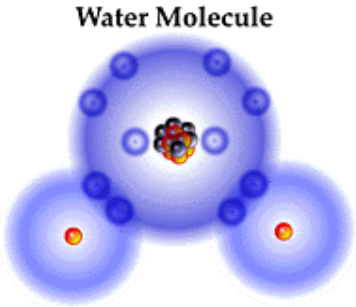
different thoughts

- 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

Putnam's twin earth

earth

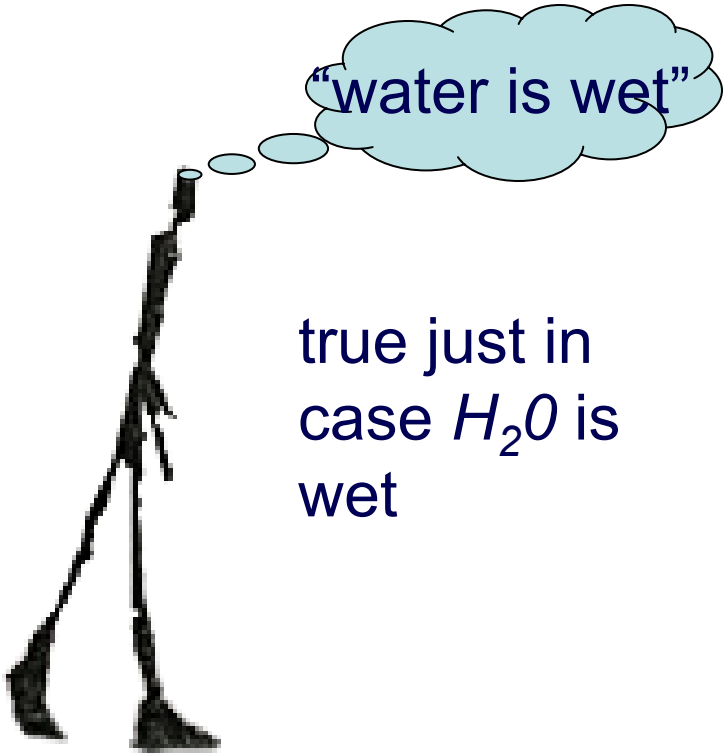
twin earth



just like earth, except that the oceans and lakes contain "XYZ", which is a very different chemical kind from H_2O, although superficially like it at normal temperatures and pressures

- let us ignore the complication that our bodies contain lots of H_2O
- further, let's pretend that no one (on earth or twin earth) knows any chemistry (accomplished in Putnam's example by "rolling the time back to about 1750")

twin-Gene singing in XYZ
on twin earth



$Oscar_1$ (on earth)

$Oscar_2$ (on twin earth)

Putnam's example seems to show that some mental properties (like the property of believing that water is wet) are not intrinsic

"Individualism and the mental"

- Putnam's example arguably shows that differences in the subject's environment (e.g. H_2O vs. XYZ) can by themselves make a mental difference
- Burge's examples purport to show that differences in the subject's *linguistic community* can by themselves make a mental difference

Burge's thought experiment

- **stage 1**
  - Alfred has various beliefs about arthritis: that he has had arthritis for years, that stiffening joints is [are] a symptom of arthritis… (all true)

    and:
  - that he has arthritis in his thigh (false, because arthritis in an inflammation of the joints)

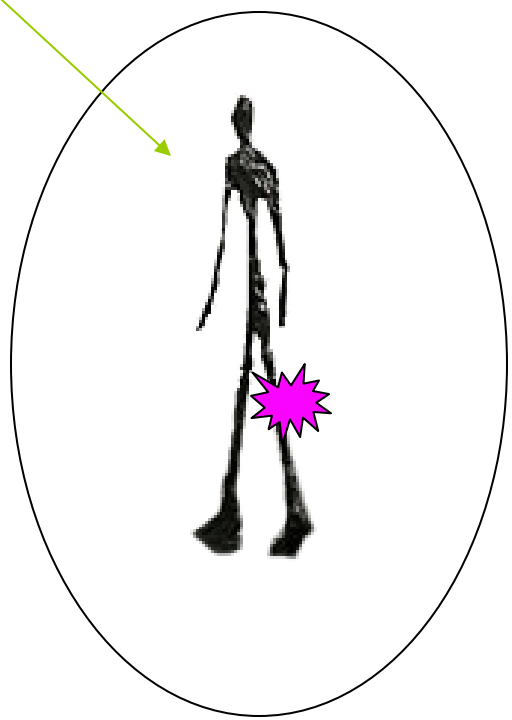
Burge's thought experiment

- **stage II**
  - a "counterfactual situation" (a non-actual possible world) in which Alfred is exactly the same in all intrinsic respects, but lives in a slightly different linguistic community
  - in this community, 'arthritis' applies "not only to arthritis, but to various other rheumatoid ailments"
  - in the language of this community, 'Alfred has arthritis in his thigh' is true

Burge's thought experiment

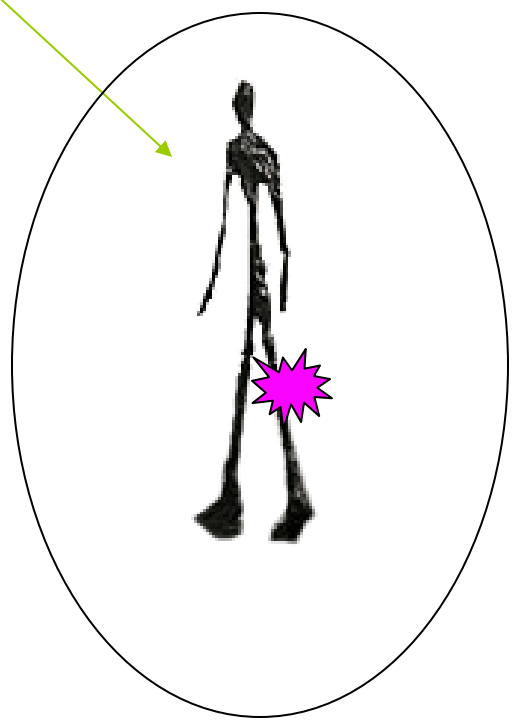
- **stage III**
  - an "interpretation of the counterfactual case"
  - Alfred has no beliefs about *arthritis* (in particular, he doesn't believe that his has arthritis in his thigh)
  - instead, he has beliefs about the sort of general rheumatoid ailment that is labeled in his community by the word 'arthritis'

Alfred with arthritis beliefs



@ (the actual world)

Alfred (a duplicate of Alfred as he is in @) without arthritis beliefs



w_1 (the counterfactual situation)

some consequences

- externalism results in a problem about *mental causation*—to be discussed next week
- externalism also results in a problem about *self-knowledge*: see McKinsey, “Anti-individualism and privileged access”
- but for now, recall our discussion of Searle’s Chinese room argument

derived/underived (or “original”) intentionality

Something has *derived intentionality* just in case it has intentionality in virtue of the intentionality of something else. Plausibly, `dog' refers to dogs in virtue of the beliefs, intentions, etc., of English speakers—hence `dog' has derived intentionality. My belief that dogs have fur is an intentional state, and doesn't have its intentionality in virtue of the intentionality of anything else—hence my belief has *underived* (or *original*) intentionality. If thinking is conducted in a language written in the brain, then the words of this language have underived intentionality.

the robot reply

"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.

weak strong AI and externalism

- our discussion of the Chinese room argument suggested that STRONG STRONG AI was false, but that WEAK STRONG AI might yet be true
- our discussion of externalism has reinforced this conclusion
- maybe if the Chinese room is hooked up in the right way to its environment, then it will have the sort of (underived) intentionality distinctive of mental states
- a suggestion about the “right way” is our next topic

Minds and Machines

spring 2003

Content: psychosemantics

“A recipe for thought”

- a sketch of a “naturalistic” account of intentional mental states (a “psychosemantics”)
- “Thought may be intentional, but that isn’t the property we are seeking a recipe to understand. As long as the intentionality we use is not itself mental, then we are as free to use intentionality in our recipe for making a mind as we are in using electrical conductors in building an amplifier or gumdrops in making cookies”

Fred Dretske

Dretske's example of "original" (underived) intentionality

the compass indicates (when used properly) the location of the north pole, not the whereabouts of the Three Bears (even if the Three Bears are at the north pole)

- so the way the compass represents seems importantly similar to how beliefs represent—one may believe that the location of the pole is over there and not believe that the location of the Three Bears is over there (even if the Three Bears are at the north pole)
- see "Intentionality" on inten**t**ionality/inten**s**ionality

> “Intentional systems, then, are not the problem. They can be picked up for a few dollars at your local hardware store.”

But:

“We are…trying to build systems that exhibit what Chisholm describes as the first mark of intentionality, the power to say that so-and-so is the case when so-and-so is not the case, the power to misrepresent how things stand in the world. Unlike compasses, these fancy items are not to be found on the shelves of hardware stores.”

misrepresentation

- when the compass is used correctly (in particular, with no magnetic interfence), the needle will always point north
- that is, without interference, if the needle points in direction *d*, then *d* is the direction of the north pole
- this fact is does not depend on the purposes and attitudes of the designers and users of compasses
- this is the sense in which the compass has underived/original intentionality: without interference, it infallibly indicates the direction of the north pole

misrepresentation

- interference is possible: a tv set might cause the needle to point east
- in this situation, the compass *misrepresents* the location of the north pole
- but: the compass only misrepresents the location of the north pole because of "the purposes and attitudes of its designers"
- so the compass doesn't help us understand how a physical system could exhibit the "first mark of intentionality"—the power to *misrepresent*

Minds and Machines

spring 2003

- read Crane on internalism and externalism
- read Dennett