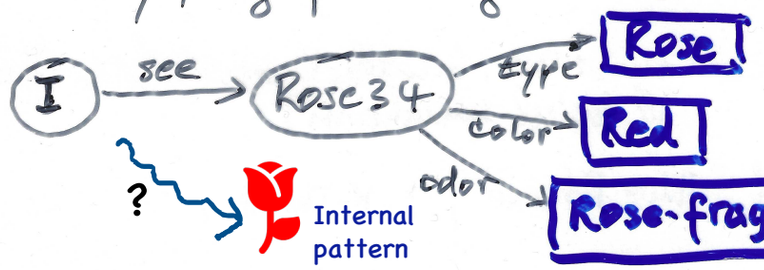
McDermott, ch.3 (&4)
Computational theory of consciousness

- Will feel like "explaining away"
- HO Theory; eg. perceiving a rose



For McDermott, "symbols" can, but don't have to, denote something

primitives of the internal self-model

- Explaining "pain in the foot" (Jackendoff's problem)
- Explaining "free will": ∞ regress in self-prediction
  [but what abt. causal explanation after the fact?]
- Pleasantness/unpleasantness & "primitive goals" [drives]
  ⇒ qualia
- Robots & emotions

  Emotion = belief + preference (- .... +) + distinguishing qualia
  to trigger appropriate reaction [e.g., ?]

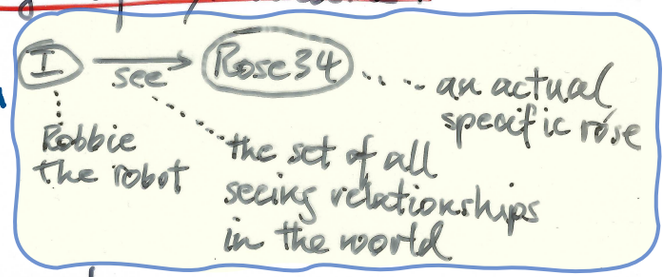
- Perceptions, & introspective awareness/judgement of these:
  "the stick in the water" - perception viewed as perception
  Also: stereo pics [& mirror images, phantom puddles, movies, ...]
- Qualia: Imagine a robot describing/comparing colors of objects ... it can't provide reasons for judgements
  So [?], it believes it experiences the primitives

[what about abstract primitives - being an object, a distance, a line, a point, a number, etc.?]

McDermott, ch. 3, cont'd

[A note on aspects of "meaning" of symbolisms:

- denotational semantics (allows a definition of truth and of sound inference)
- operational semantics: how the symbolism enables/supports cognitive processes & action
- conceptual semantics: How symbolic representations are structured & related to one another (often, how complex conceptual representations are built up out of a set of primitives)
- causal semantics: How internal symbolic representations are "caused" by interaction with the world (perception, verbal communication), & in turn how the representations cause the agent to act on the world.]



- McDermott: Representation as "resemblance"
  [but this seems more true for analog representations, eg. diagrams, than purely symbolic ones (see above)]
- Self-model: crucial to consciousness

[Confusing discussion of "I" (not worth trying to sort out, IMHO)

Literal quotes or scare quotes?

"I" — symbol for the agent? "I" is the creature who makes decisions
"I" — an active "module"? "I" is an object of the self-model

p 123: "Robot 1 believes Robot 2 to have (or be) a self like its own 'I'." [?? A symbol is not a "self"]

McDermott, ch. 3, cont'd + ch. 4

Self-model, cont'd

- Self-fulfilling self-ascribed properties: e.g., having certain intentions. [But from a Baars / Libet perspective, conscious intentions are high-level abstractions "posted" by prior unconscious processing.]
- Location: Apparently in left hemisphere, near speech centers (Gazzaniga)
- Autobiographical (episodic) memory "seems to call for a general internal representation" [ok!] ... but this is "far-fetched" [why??]
- "Being cognizant of" = "self-model having access to a representation of"
- p126-130: Thought experiments about language separate from consciousnes [hmm...]

ch. 4: Objections & Replies

- Argues against "raw feels" without a representation in the self-model (or accessible to it?) (contra Ned Block, Michael Tye, Larry Shapiro, ...)

  Shapiro: it's the tickling/barking that I experience, not my awareness of these sensations!

  McD: Sure, but for the sensations to be conscious they must be represented in the self-model

  [What about "absorption"? How does accessibility "turn on" consciousness?]

McDermott, ch. 4, cont'd

- Qualia (reiterating): "A quale is nothing but the brain's way of thinking about its own sensory comparison system" [so, qualia = 2nd order thought?]

- Computers – why aren't they conscious yet?
  Ans.: Impoverished mental world [babies?]

(See more "picturesque" versions, upcoming slides)

- Zombies (absent-qualia argument): McD. imagines Chalmers' zombie double arguing it (unlike its supposed double) is conscious!
- Mary, the color-deprived color theorist (Frank Johnson's disproof of materialism)
  McD.: Mary will say "Now I know red looks like this" but "this" is gibberish from an external perspective – an arbitrary internal symbol – better: pattern
  [But surely she does learn something, subjectively?!]
- Inverted spectra (twins)
  McD.: Intersubjective comparisons are meaningless
  [yet I suspect your blue/pain/etc. are much like mine!]
- What's it like to be a bat? (T. Nagel)
  McD.: same as for inverted spectra...
- Searle's Chinese Room (neither the person nor the rule book have any awareness of what's going on...)
  McD.: The system as a whole does... [if fast enough?]
- Block's billion-people brain simulator
  McD.: Intuition should yield to reason
- Qualms about brain as confederacy [self ⇔ self-model confusion?]

Some fun topics concerning phenomenal consciousness (see McDermott ch.4)

The real David Chalmers:

I experience qualia, but conceivably, I could have a double who is just like me
but experiences nothing!

Zombie-universe Chalmers:

I experience qualia, but conceivably, I could have a double who is just like me
but experiences nothing!

Key considerations:

Partial zombies …

- Captioning glasses
- Sleepwalkers
- Blindsight
- Cognitive robots

Mary, the color perception expert:
(Frank Johnsons' thought experiment)

I've grown up in a black-and-white lab, but I fully understand how my brain would process the sight of a red rose!

Wow, I didn't expect that!

I've learned something new!

(Therefore qualia are not explained
by neuroscience.)

Key considerations:

- Purely symbolic self-modeling
- Self-models with "signal sites"

BTW, McDermott invokes something like "signal sites", but says they are "gibberish" from an external, linguistic perspective.

Inverted spectra

Logically, could this be?

If so, are qualia non-causal?

Key considerations:

- Near-identical brains => near-indentical qualia??
- What about less similar brains?
- What about robots, bees, aliens?
- Relevance of qualia topology

John Searle's "Chinese Room"

Searle knows no Chinese, yet "converses"

Are computers like this, & thus unconscious, & unintelligent?

If you see this shape,
"什麼"
followed by this shape,
"帶來"
followed by this shape,
"快樂"
then produce this shape,
"爲天"
followed by this shape,
"下式".

Key considerations:

- Might the "system" of {room + stored information + Searle} be conscious & intelligent?
- Would the required storage fit into the universe?
- Does Di Li's (MS Asia) hugely popular "Xiaoice" implement Searle's idea?

For dialog examples see http://nautil.us/issue/33/attraction/your-next-new-best-friend-might-be-a-robot