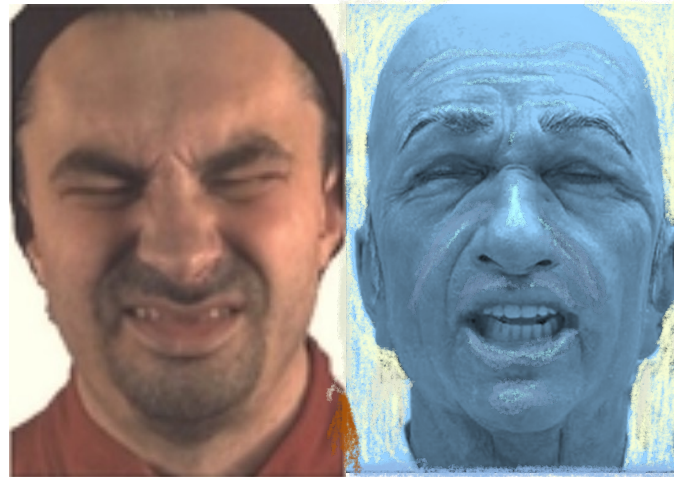


CSC 191/291/21 Machines and Consciousness




Self-awareness (“Know Thyself” by unnamed artist (owner Dovanna Cleary), Salvador Dali, and Art Levin)



Phenomenal Consciousness (Colors, pains, etc.; dreams)

Introduction

Len Schubert,

WH 3003 (when life goes back to normal ) ,
<http://www.cs.rochester.edu/~schubert/>

Grad TA (1/2) Lane Lawley

UGs Elana Elman, Catherine Giugno (?)

191 prerequisites: a programming course; sophomore or higher standing (or consent of instructor); a prior course in AI, cognitive science, linguistics, or philosophy is an advantage.

291 prerequisites: 173, one of 200, 240, 242, 212 (HCI), 249 (computer vision), 260 (topics in NLU) (or perhaps a theory course -- 280 or 282).

Office hours: LS: Tue. and Thu. 5:00–6:30 pm, **TAs:** TBD

Materials & logistics

Texts: (covered selectively, & supplemented with some articles, handouts)

McDermott, *Mind & Mechanism*

Blackmore, *Consciousness*,

Baars, *I the Theater of Consciousness*,

Fellous & Arbib, *Who Needs Emotions*

% weight 191/291

- | | | |
|---|----|----|
| - approx. 5 homework assignments (answer questions, write essays) | 40 | 35 |
| - 2 staged Lisp exercises (oriented towards "self-awareness") | 26 | 16 |
| - midterm test and noncumulative final test | 34 | 24 |
| - Gridworld projects (teams of 1-3) | | 25 |

Course description and tentative schedule:

See CSC 191/291 instructor and TA website. (NB: homeworks, test schedule)

Reading assignment for Thursday (see 'some lecture materials' link on my website):

Steven Pinker, "Could a computer ever be conscious?" (a talk)

[Cory, Alicia, & Allenby; "raw feels" vs self-monitoring, independence of materials used, duplication, matter-transporter; can you murder a robot, inflict pain on a baby or a lobster or a worm? We won't know for quite some time.]

Goals of the course

- Know how to **represent basic self-knowledge** in a machine, and enable QA
- (CSC 291 students) Develop **skills in self-aware agent design**, symbolic programming
- Understand (as far as current knowledge allows) **the meaning of "consciousness"** (both *access consciousness*, and "*phenomenal consciousness*", including self-awareness)
- Develop an informed, **reasoned view of the basic issues** concerning consciousness, in humans, animals and robots (from philosophical, cognitive, and AI perspectives)
- Acquire basic understanding of the **biological basis of consciousness** and emotion in people (& animals)
- Acquire basic understanding of **AI/robotics efforts to endow machines with consciousness and emotions**
- Hone **skills in philosophical and scientific writing/argumentation**
- **Stir up interest** in one of the most intriguing topics of all time

Introduction

- Consciousness and the taboo of subjectivity in 20th century (“unscientific”)
- Since 1980s: clear connection between verbal reports and brain activity; AI cognitive engineering perspective
- Will you see truly conscious, intelligent machines in your lifetimes?
- Well, it depends (human self-destructiveness, research fashions)
If so, what are implications for ownership, autonomy, ethics, laws, and humanity's future? Cf. Corry & Alicia (or Blade Runner, AI, Ex Machina, etc.)

Kinds of consciousness

- We're aware of, and can report on, our own characteristics, history, relationships, possessions, etc. → “access consciousness”
- We have sensory impressions/experiences – colors, shapes, motions, sounds, fragrances, pains, pleasures, emotions, real & imaginary → phenomenal consciousness (qualia)

How do these work? We'll try a “cognitive engineering” perspective – which has advantages over biological/philosophical ones!

Self-awareness & self-knowledge

- The ambiguity of "*self-awareness*"
- We'll focus initially on (reportable) *self-knowledge*
- An excerpt from Bertrand Russell's autobiography ...
- Could a machine have this kind of reportable self-knowledge?
We'll need a little bit of CSC 244 & 247 to address this;
We'll also need to learn Lisp – still the best PL for symbolic processing