

Fellous & Arbib



Michael Arbib,
USC, UCSD

ch. 12, Arbib - Beware the passionate robot

- Robots, unlike Arbib himself, shouldn't have fits of pique, leading to rudeness & possible dire consequences
- Function of emotion: biasing toward rapid commitment to action (vs. more cautious appraisal)
 neuromodulation (dopamine, serotonin, opioid peptides)
- Our most powerful emotions are rooted in basic drives (the "4 F's"), rather than expected/learned rewards & punishments. Such "heated emotions" may be inappropriate for robots.
 ONR (ch. 7) leave out the "heat"; Rolls too focused on +reward -punish

Consciousness & Language

- Full human-like consciousness depends on language [definitely seems true for consciousness of our thoughts - inner speech]

Evolution of (vision and) language

visual features → action (moving dot → feed; looming obj → flee)

↓
action-neutral representations,
visual action monitoring (hands)

dorsal: affordances
ventral: concepts

↓
triggering neural circuits for monitored action
when observing other primates doing the action

mirror
neurons
(Opp. Broca)

↓
pantomime to convey intended action

↓
pantomime + vocalization ⇒ full, recursive language
& protosign (symbolic)

Arbib, ch. 12, cont'd

(2)

- objects to Fellous & Ledoux's claim that consciousness \approx working memory (also Baars?)
- objects to Rolls' hypothesis that monitoring of "if-then" rules is dependent on language
- empathy: We comprehend other minds by simulation \Rightarrow empathic emotional response [except in sociopaths? Marie Bresnard...]
- world graph model of the rat (with Liellick '77) [Maze-running model?] current-situation model, in prefrontal cortex; but "where am I" (current spatial context) is represented in the hippocampus [cf. London taxi drivers]
- Robots will have "functional niches" & their dispositions & reward systems should be adapted to these.
incidental - rats & lever pressing (hypothalamic electrodes) "wireheading"
[would robots manipulate their own reward variables?!]

The opening story showed that for people, emotions can have bad consequences (cf. massacres in the news...). So we don't want that kind of "heated" emotion in robots... restrict them to "cautious appraisals".

Arbib urges a goal-oriented design, balancing the need for rapid commitment to action with more careful evaluation