Transhuman intelligences will be hyperemotional

Claudius Gros

Institute for Theoretical Physics Goethe University Frankfurt, Germany



synthetic intelligences_

from

- algorithmic problem solvers
- Deep Blue, Watson, AlphaGo, autonomous driving, . . .

to

- artificial (trans-human) intelligences
- autonomously planning
- self motivated



the motivational problem ____

standard view advanced AIs as evolved utility maximizers

long-term utility maximization possible?

utility function

unknown does one exist? not computable scarce resources

information computational power time

nature's solution

- diffusive emotional control
- logic useful to achieve goals efficiently, not to set them

neurotransmitters _

local trans-synaptic chemical information transmission



GABA : inhibitory glutamate : excitatory

neuromodulators

modulating

- synaptic plasticity neural thresholds, gains, ...
- * norepinephrine
- dopamine
- ★ serotonin
- * choline, oxytocin, ...



[Physiological Reviews]

no direct cognitive information processing - diffusive control

neuromodulation & cognition _

[Emotions: from brain to robot. Arbib, Fellous, '04]



» higher cognition essentially dependent on neuromodulation «

moods, emotions and neuromodulators ____



- most neuromodulation is neutral (unconscious)
- why do we experience certain neuromodulatory processes as emotions?

James-Lange feedback



"sensation" of emotions through feedback (body)

neutral & emotional neuromodulation _____



why has nature developed emotional neuromodulation?

what activates neuromodulation? _____

external and internal sensation

- C. Elegans (nematode, 302 neurons)
- Drosophila (fruitfly)



cognitive processing (thoughts)
 – couch experiment

control of neuromodulation ____

neutral neuromodulation $\hat{=}$ open loop control

hungry \rightarrow increased response to food stimuli

 \rightarrow food uptake

 \rightarrow fed

emotional neuromodulation $\hat{=}$ closed loop cognitive control

- angry \rightarrow James-Lange bodily response
 - $\rightarrow\$ cognitive awareness of mood
 - \rightarrow cognitive control

note: a defendant is liable (in court) for its action when capable of closed-loop cognitive control of her/his motives/emotions

emotions & motivational problem _

goal setting

comparative evaluation of (emotionally attributed) values

value attribution to action options

- genes (simple options)
- acquired (advanced cognitive systems)
 e.g. preference to play guitar instead of violin

open loop control (evolution)

play guitar (why?), evaluate consequences

closed loop control (emotions)

think about playing guitar (using memories, ...)

cognitive pyramids_

present-day AI roadmap

emotions consciousness

learning, problem selection

vast bank of specialized algorithms

biological intelligences

knowledge, abstract reasoning, specialization

consciousnes, emotions learning

> neuromodulation genes

increasing intelligence _

- opens behavioral options
- needs advanced comparative weighting mechanisms
- sophisticated closed loop emotional control system

transhuman intelligences will be hyperemotional

regarding sophistication, not intensity