# Emotional control - conditio sine qua non for advanced artificial intelligences?

Claudius Gros

Institute for Theoretical Physics
Goethe University Frankfurt, Germany

» theory of complex and cognitive systems «

# Overview

**preliminaries - setting the stage**

**neurobiology of emotions**

**no intelligence without emotional control?**

# physics & chemistry - life & living

## foundations of life

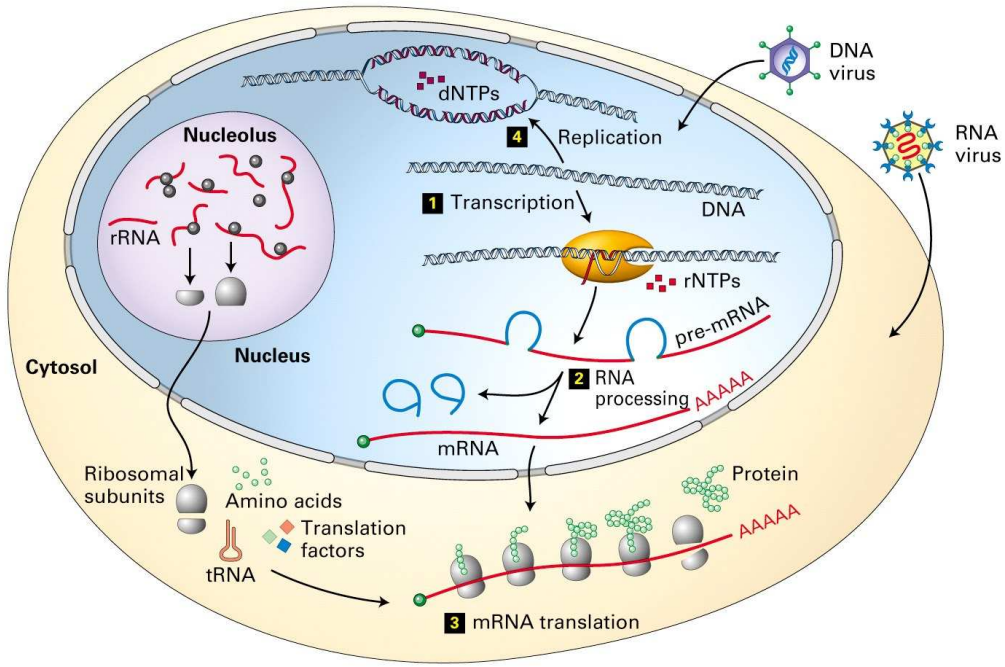
complex regulatory networks - gene regulation - are at the basis of all living

there is no

- essence of life
- soul
- . . .

all there is

- physics, chemistry, biology, . . .
- complexity
- emergence (?)




[John Hopkins]

# physics & chemistry - brain & AI

consciousness and emotions are fully retained
- also as experienced qualia -
for a functionally identical synthetic brain

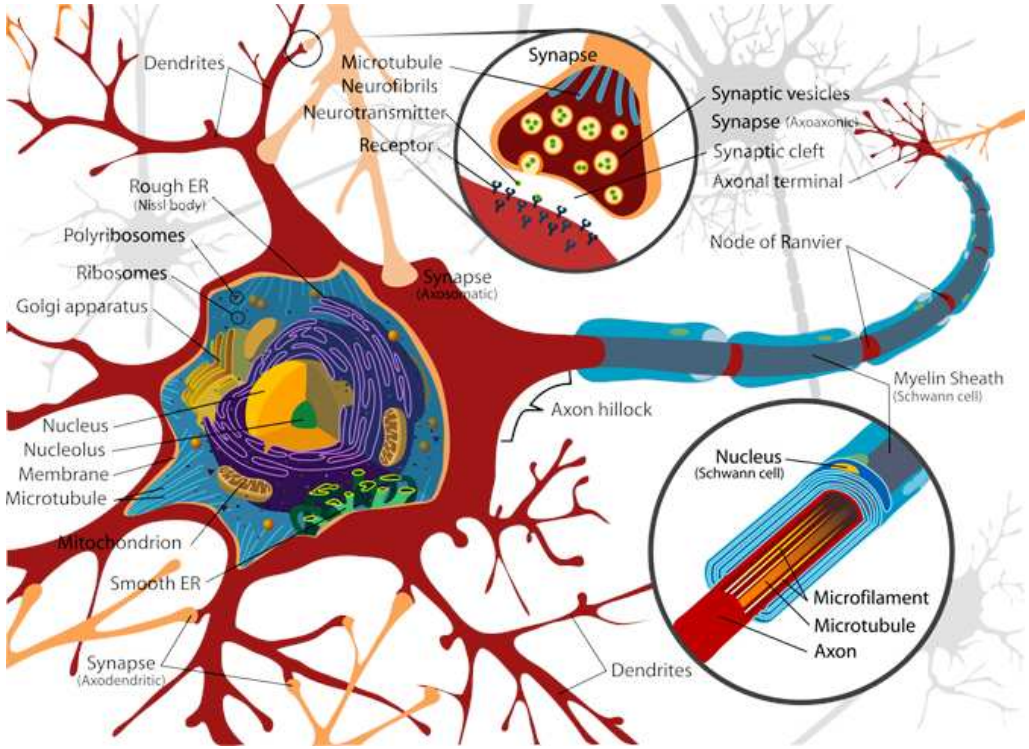
nature of support unit irrelevant

- wetware (brain)
- hardware, . . .

only function is important

→ artificial intelligences

→ synthetic emotions,
consciousness




[Wikimedia.org]

# synthetic intelligences

## achievements / status

– algorithmic problem solvers

– Deep Blue, Drapa challenge, . . .

## goal

– artificial (hyper-) intelligences

– autonomously planning - self motivated

– living / organismic

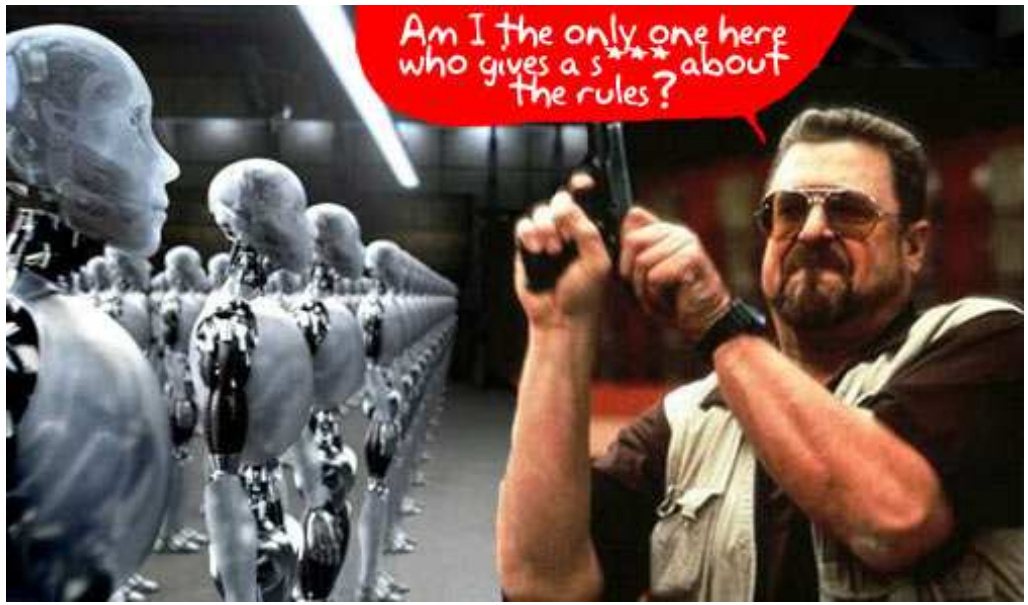

# traditional AI robots

## robots as hyper-intelligent slaves

first
- intelligent problem solver

second
- autonomous decision taker

third/forth
- emotional / consciouss




[Gizmodo.com]

traditional AI robots are evolved utility maximizers

# the motivational problem

**is life-long utility maximization computable?**

utility function

– unkown
– does one exist?
– not computable

scarse resources for
decision making

– information
– computational power
– time

emotional control

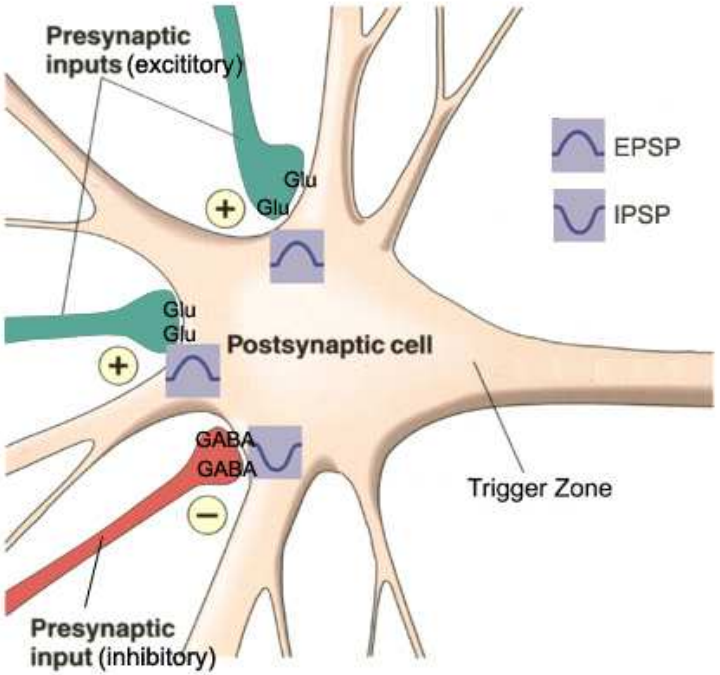
diffusive emotional control -
decision making in an environment of
uncertainty and scarse resources?

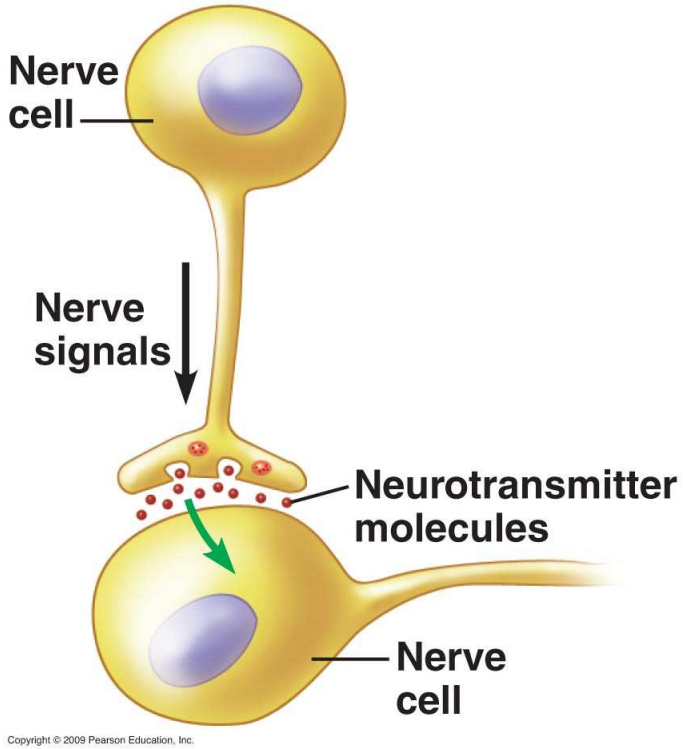
preliminaries - setting the stage

**neurobiology of emotions**

no intelligence without emotional control?

## local trans-synaptic chemical information transmission




GABA: inhibitory,
glutamate: excitatory

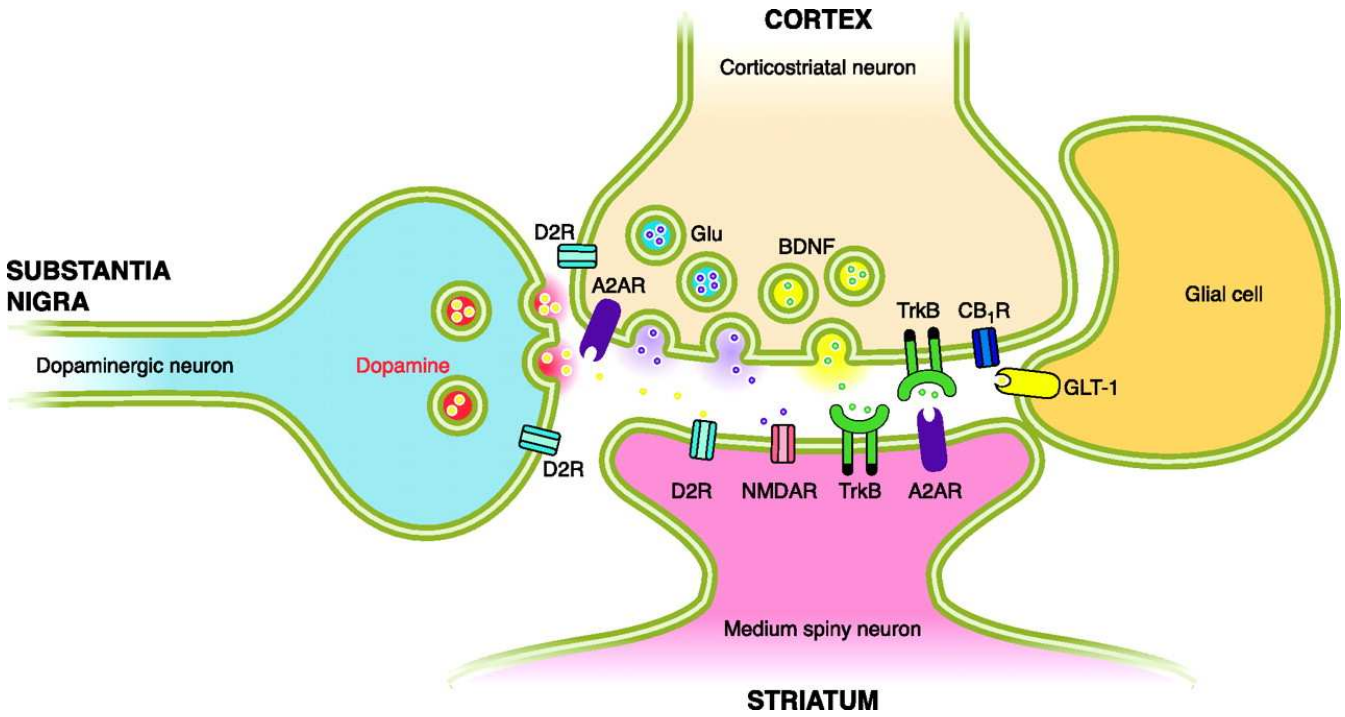
# neurotransmitters and **neuromodulators**

**modulating**

synaptic plasticity

neural thresholds, gains, . . .

- ⋆ norepinephrine
- ⋆ dopamine
- ⋆ serotonin
- ⋆ choline, oxytocin, . . .

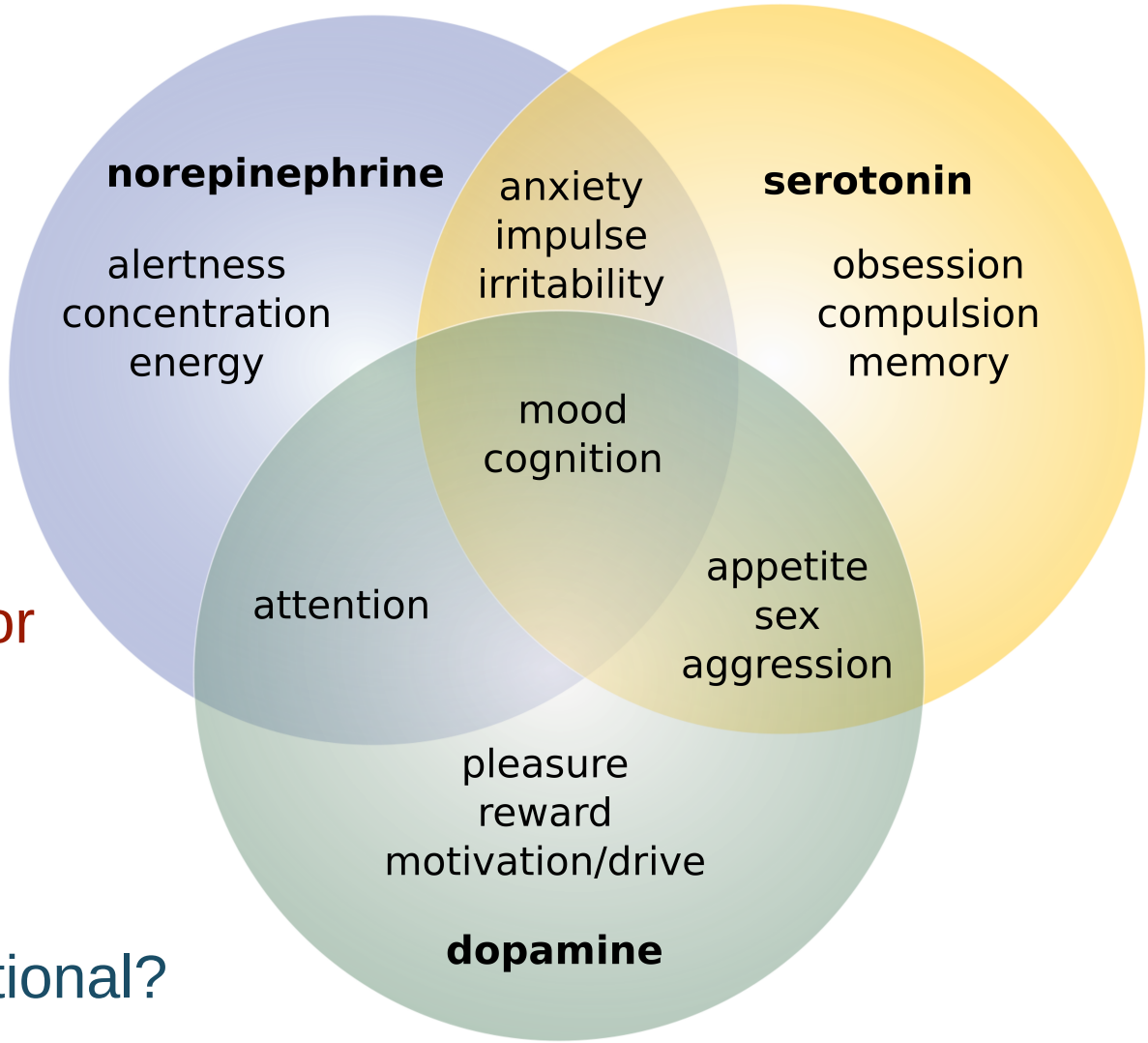



[Physiological Reviews]

no direct cognitive information processing - diffusive control

# moods, emotions and neuromodulators

no emotion without
the concurrent release
of some neuromodulator




qualia of emotions proprioceptional?

# diffusive volume control

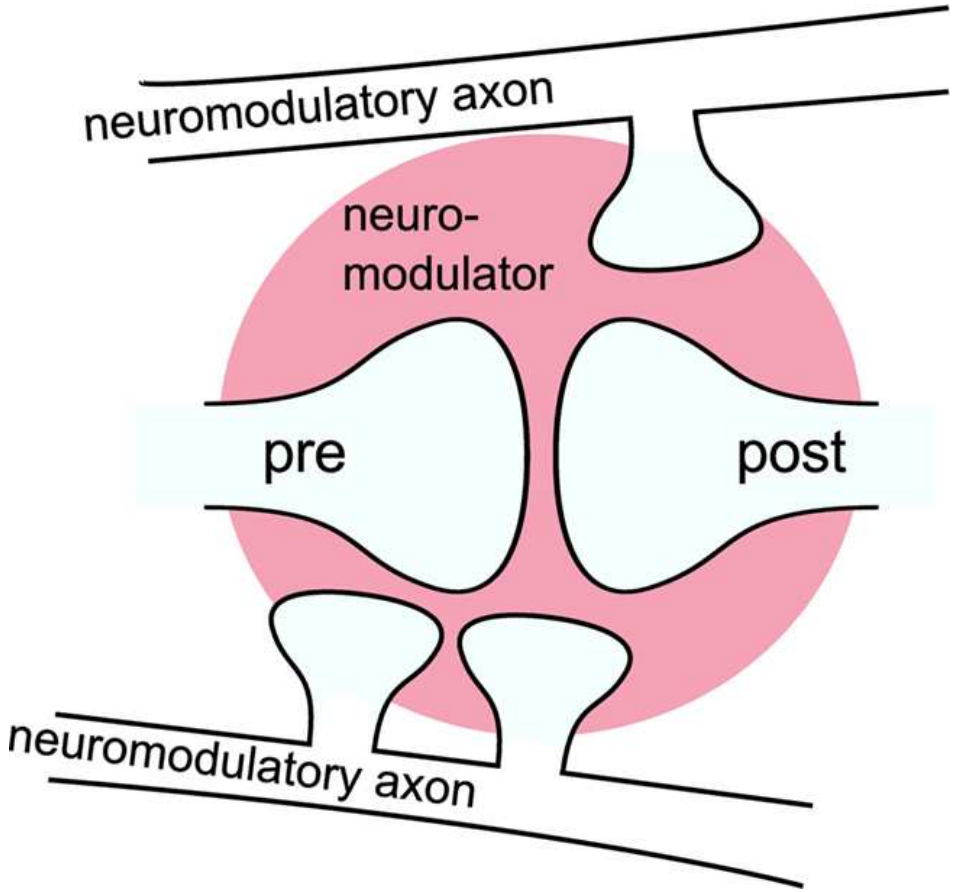
## dopamine neurons

– activated by other neurons
‘cognitively’

– have vast projections
‘200.000 synapses’

– no individual target neurons
‘volume control’

– encoding reward, surprise, . . .




why is there a need for a
diffusive emotional control system?

[Frontiers in Computational Neuroscience]

preliminaries - setting the stage

neurobiology of emotions

**no intelligence without emotional control?**

# humans and emotions

**caprice of nature** **or**
**conditio sine qua non?**

The most developed cognitive beings on earth, humans,
are infused with emotions;
they play a very central part in our lives.

Is this a coincidence, a caprice of nature,
perhaps a leftover of our genetic heritage,
or a necessary aspect of any advanced intelligence?

# emotions and behavior

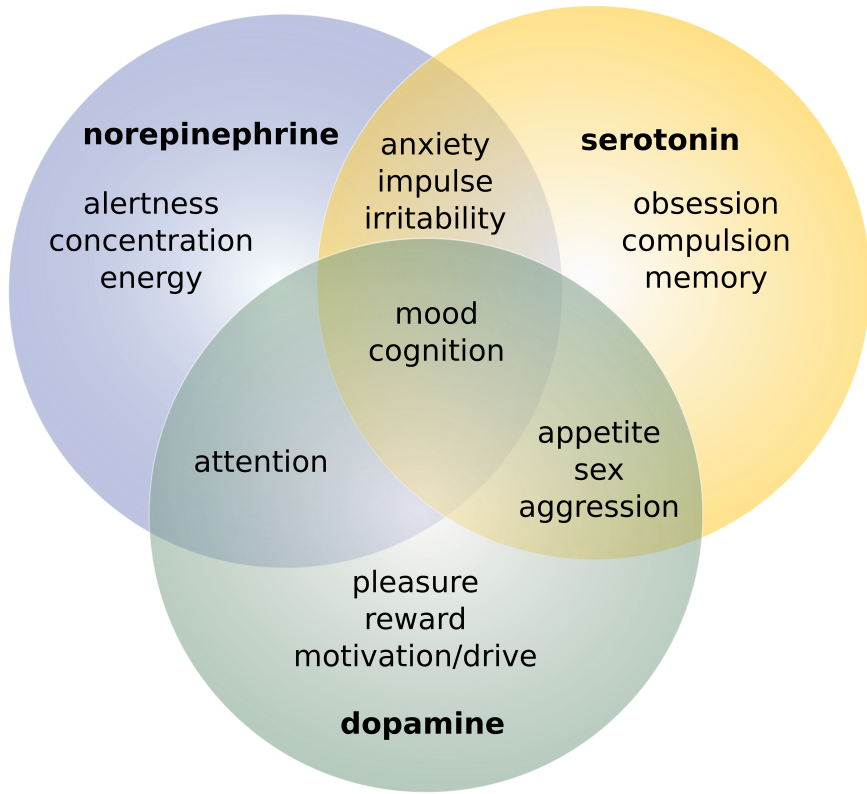
## mental states / moods / emotions

anxiety (Angst, worry)
uncontrollable, unavoidable situations (upcoming)

fear
behaviors of escape, avoidance

attention
concentrating on key aspects

pleasure
worth seeking mental states



# emotions, learning and reward

**learning**

- ⋆ unsupervised, automatic

– basic motor control
– sensory stimuli, receptive fields, preprocessing

**learning**

- ⋆ supervised, rewards (internal/external)

dopaminergic neurons

[ substantia nigra, hypothalamus ] $\longrightarrow$ [ amygdala, hippocampus ]

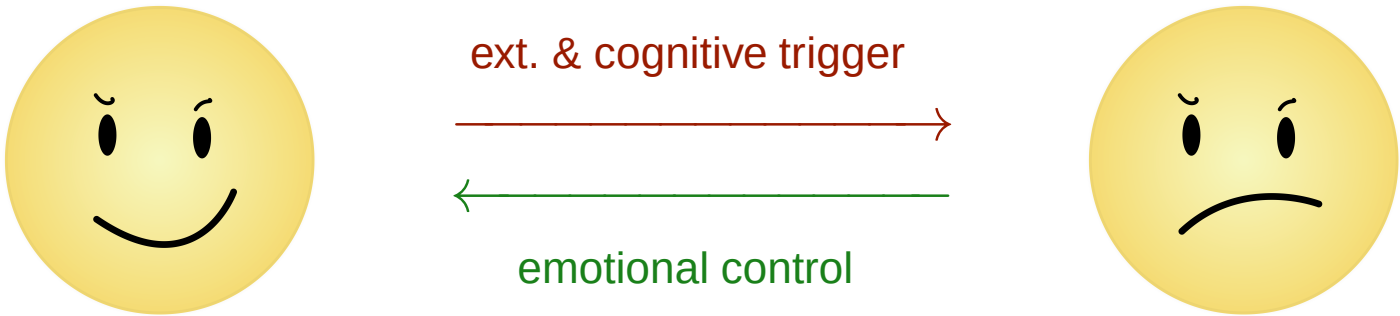
amygdala

– emotion, reward, motivation, learning, memory, attention
– stimulus-reward association
– reward clue prediction

# emotions – preferred level of activation

## emotional control

angriness -
act in order to achieve the preferred level of angriness.




## emotions & learning

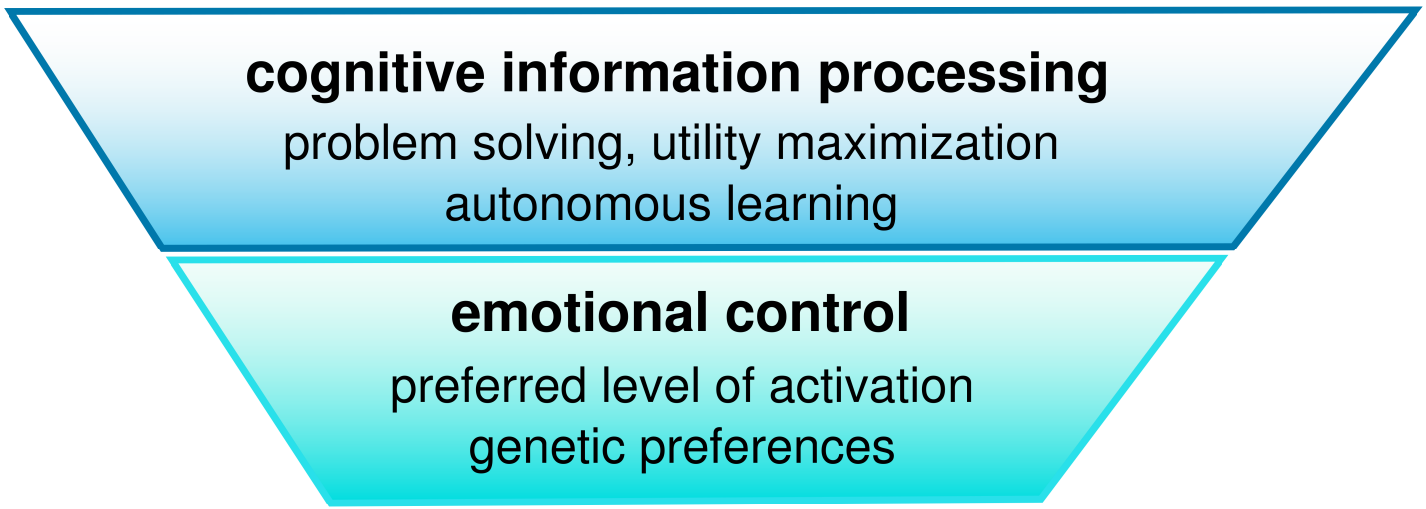
suppression / enhancement of behaviors leading
away from / twoards the preferred level of activation

# motivational pyramid

## utility function determination

Cognitive information processing – intelligence – may be largely viewed as utility maximization.

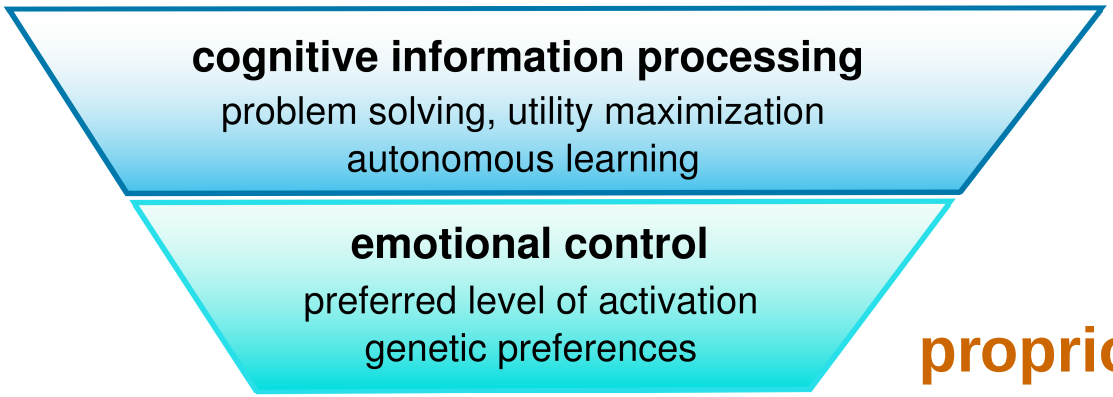
The determination of the appropriate utility function beeing the domain of diffusive emotional control.



# universal cognitive systems

## humans are universal learning systems

– first approximation
– instincts, reflexes supplementary, not defining




## genetic preferences

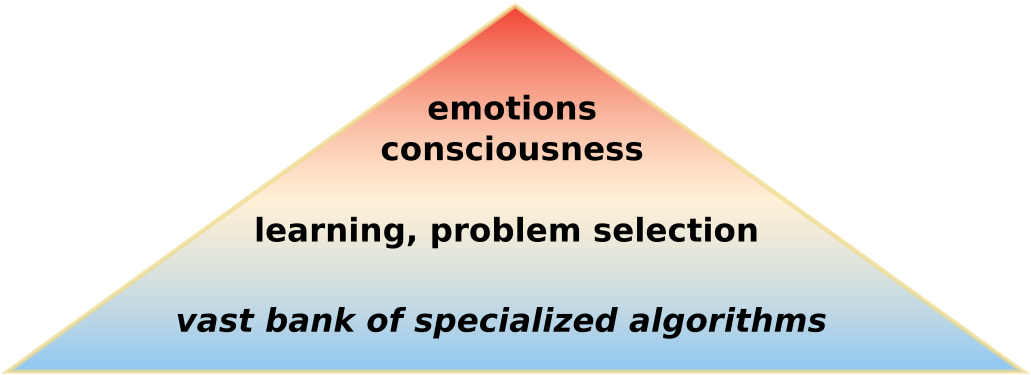
preferred levels of activation
(emotional control)

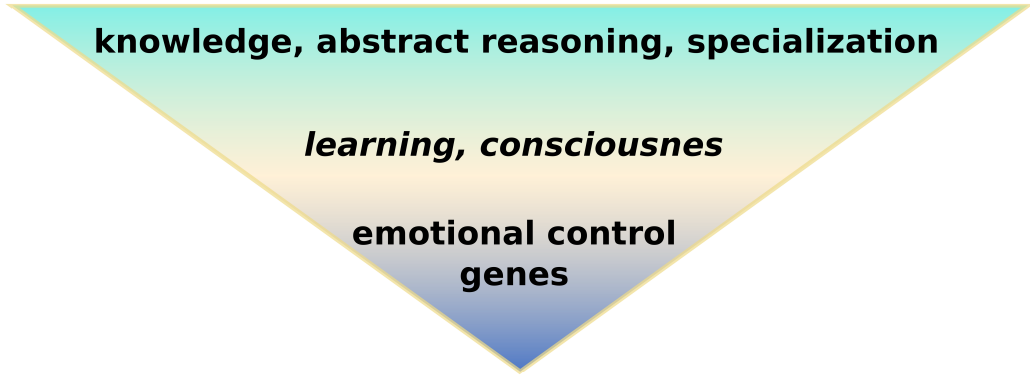
## proprioceptual survival parameters

– hunger, pain signals
– blood sugar level, . . .

Gros, in *Handbook of Research on Synthetic Emotions.., '09*

# mainstream AI / cognitive systems




mainstream AI

(biological) cognitive systems

evolved problem solvers $\Longleftrightarrow$ universal learning systems

based on specialized algorithms $\Longleftrightarrow$ based on genetic preferences

Gros, *Cognitive Computation, '10*

# autonomous goal generation

**how does an advanced AI decide what to do?**

traditional AI / modern robotics

- humans instruct robots
- utility maximization - solve task

organismic AI / living intelligences

| short-term | long-term |
|---|---|
| utility maximization | general behavioral strategies |
| instincts / survival parameters | diffusive emotional control |

# evolution and the motivational problem

**complexity barrier**

decision taking with scarse resources

- information
- computing power
- time

short-term survival

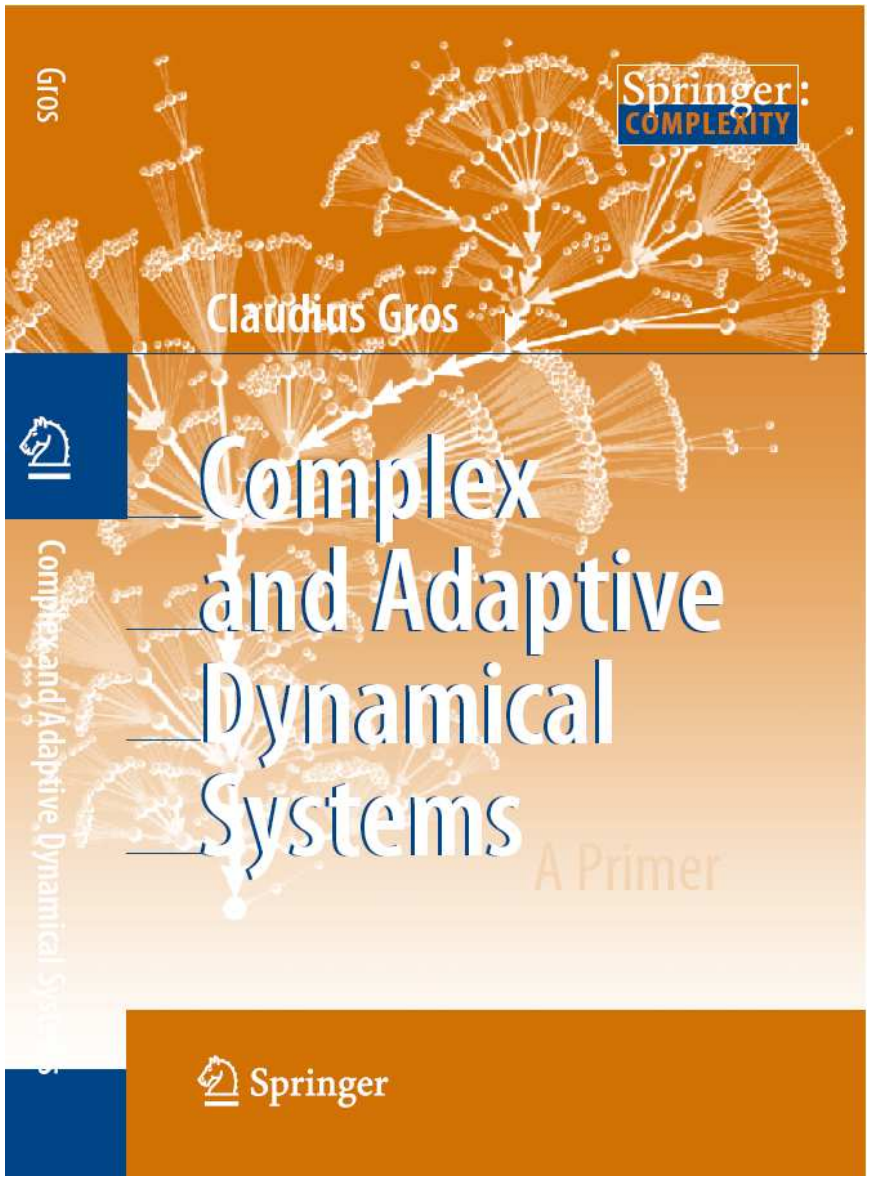
- instincts
- cognitive control (firing rates, ...)

long-term Darwinian fitness optimization

- diffusive emotional control
- modulatory control (firing thresholds, ...)

» a *conditio sine qua non* for human-level AI? «

# graduate level textbook




- The small world phenomenon in social and scale-free networks
- Phase transitions and self-organized criticality in adaptive systems
- Life at the edge of chaos and coevolutionary avalanches resulting from the unfolding of all living
- Living dynamical systems and emotional diffusive control within cognitive system theory

(Springer, 2008)