

Affect, emotion and robots

Lola Cañamero

Reader in Adaptive Systems

Head of the Embodied Emotion, Cognition and (Inter-)Action Lab

School of Computer Science

University of Hertfordshire

<http://homepages.feis.herts.ac.uk/~comqlc/>

www.emotion-modeling.info

L.Canamero@herts.ac.uk

In my research, two main preoccupations:

- **Understanding** emotions in biological/artificial systems. We ask questions such as:
 - What are the key features of human and animal emotions that can be shared by autonomous robots in similar environments?
 - How do emotions affect behavior and cognition?
 - How did emotions evolve?
 - How do they develop through life?
- **Building** better robots. We ask questions such as:
 - How can emotions help robots to make "good" decisions?
 - How can emotions make robots interact with people better?

Enhancing technology with emotions

- **Recognizing & expressing emotion: HCI, HHI through computers**
 - How:
 - adapt to user profile & current state (more sensible, adapted interaction)
 - knowing when it is appropriate to interact (not to bother!!)
 - Ex: assistants, interfaces, expressive e-mail, operating systems
- **Expressing & eliciting affect: HCI, entertainment, arts**
 - How:
 - hide the machine from the loop: natural, human-like output, believability
 - grasp & convey affect-related information (tacit): support & educational tools
 - Ex: generation of expressive music, games, interactive characters
- **Generating emotional behavior: behavior synthesis**
 - How:
 - coherence (relation with personality), conveying intention
 - persistence (for how long pursue a goal, when to interrupt), adaptation
 - Ex: animated characters, virtual worlds, robots

Affective computing (Picard 1997)

- Building artificial systems that:
 - Recognize
 - Express
 - Respond to
 - Influence
 - Facilitate
 - ***Have***
- Do “emotional” artifacts need all these capabilities?
- Do “emotional” artifacts need to *have* emotion?
- Is “strong” emotion synthesis necessary to have emotional interaction?
- *In robotics => Architectures* to generate emotions and their observable manifestations



(“human-like”?) emotion

The notion of architecture (controller)

- Main goal of autonomous agent research:
 - > come up with architectures resulting in the agent showing *adaptive*, *robust* and *efficient* behavior
- **Architecture:** set of tools, algorithms and techniques that support principles for achieving this
 - E.g., action selection architecture: *repertoire of behaviors* + *arbitration mechanism* + *design principles*
- **No optimal architecture:**
 - Understanding which architectures are the most simple solution for a given class of agent problems
 - Problem class defined in terms of the characteristics of the agent's resources (memory, sensors, etc.) and those of the task and environment

Main issues in affective computing

- **Synthesis**
 - > *enhance adaptation to environment*
 - > *persistence, coherence (relation with personality), conveying intention*
 - generation of “human-like” emotions:
 - cognitive models, “reasoners”
 - physiological, behavioral models
 - “emergent” emotions (observer’s perspective)
- **Recognition and expression**
 - > *more natural communication and interaction*
 - recognition of human emotions
 - difficulty: “continuum”, laboratory-induced emotions
 - facial expression, physiological patterns, vocal expression
 - generation of emotional expression
 - difficulty: believability (rel. believable agents)
 - movement patterns, facial expression

Emotions or Affect?

- “Wanting” and “liking”
 - Motivation for action (and for cognition, interaction)
 - Liking or disliking what we do, perceive, know, learn, how we do it, the consequences of it.
- Affect and cognition-action deeply intertwined
 - Affect drives and colors cognition-action
 - Cognition-action are part of affect
- Embodied Artificial Intelligence
 - Embodied affect

Motivational states

- Internal needs (drives, survival)
- Homeostatic model
- Vary with deprivation
- On/off -> arousal/satiation
- Stimulus-specific
- Mostly individual-centered (alertness, attention, behavior ...)
- Main functions:
 - direct behavior toward/away from specific goal
 - increase alertness
 - organize behavior into coherent, goal-oriented sequences

Emotional states

- Three components:
 - recognition of important event
 - autonomic and visceral responses
 - conscious experience in cortex
- Activation -> changes in neural & endocrine activity wrt baseline
- Generality (object, time, ...)
- Multi-faceted complex process:
 - > physiological, behavioral, cognitive-evaluative, communicative-expressive (social), subjective experience
- Main functions:
 - adaptive processes, regulation
 - categorization, evaluation
 - signals (interaction, communication)

Components of “human” emotions (Picard97)

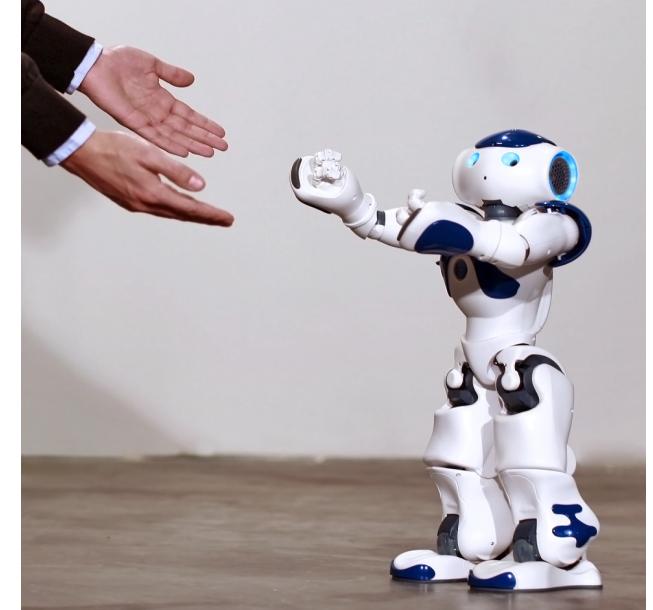
1. Behavior that seems to arise from emotions
2. Fast, “primary” emotional responses to certain stimuli
3. Ability to generate emotions from cognitions
4. Emotional experience:
 - Awareness of physiological activity
 - Awareness of cognitive activity
 - Subjective feelings
5. Emotions interact with other processes simulating human cognitive and physical functions: memory, perception, decision making, learning, goals and motivations, attention and interest, priorities, planning, immune system ...

Affective phenomena

- Motivations
 - Emotional states
 - Feelings
 - Passions
 - Temperaments
 - Moods
 - Personalities
- > all related, but different!

Embodied affect

- Much more than using physical artifact
- More than expression and recognition
- Embodiment beyond “shape”:
 - Affect happens *in* and *through* the body (fundamentally embodied)
 - “Internally” (e.g. “hormones”) and “externally” (facial and bodily expression, other behavior, interaction)
- To what extent can similar affective mechanisms be used across embodiments?
- Humanoid embodiment?
 - Anthropomorphism ... good! ... bad!
- Human-like affect for robots?



What are emotions?

- Difficult to define (and identify!)
- Many problems :
 - > e.g., discrete categories or continuum?
- Multi-faceted phenomena:
physiological, behavioral, cognitive, social ...
- Three main elements (development & activation):
 - Recognition of significant event
 - Autonomic and visceral responses
 - Emotional experience in cortex
- Which are the causal relationships among the 3 elements?
 - > different theories

Approaches to human emotion

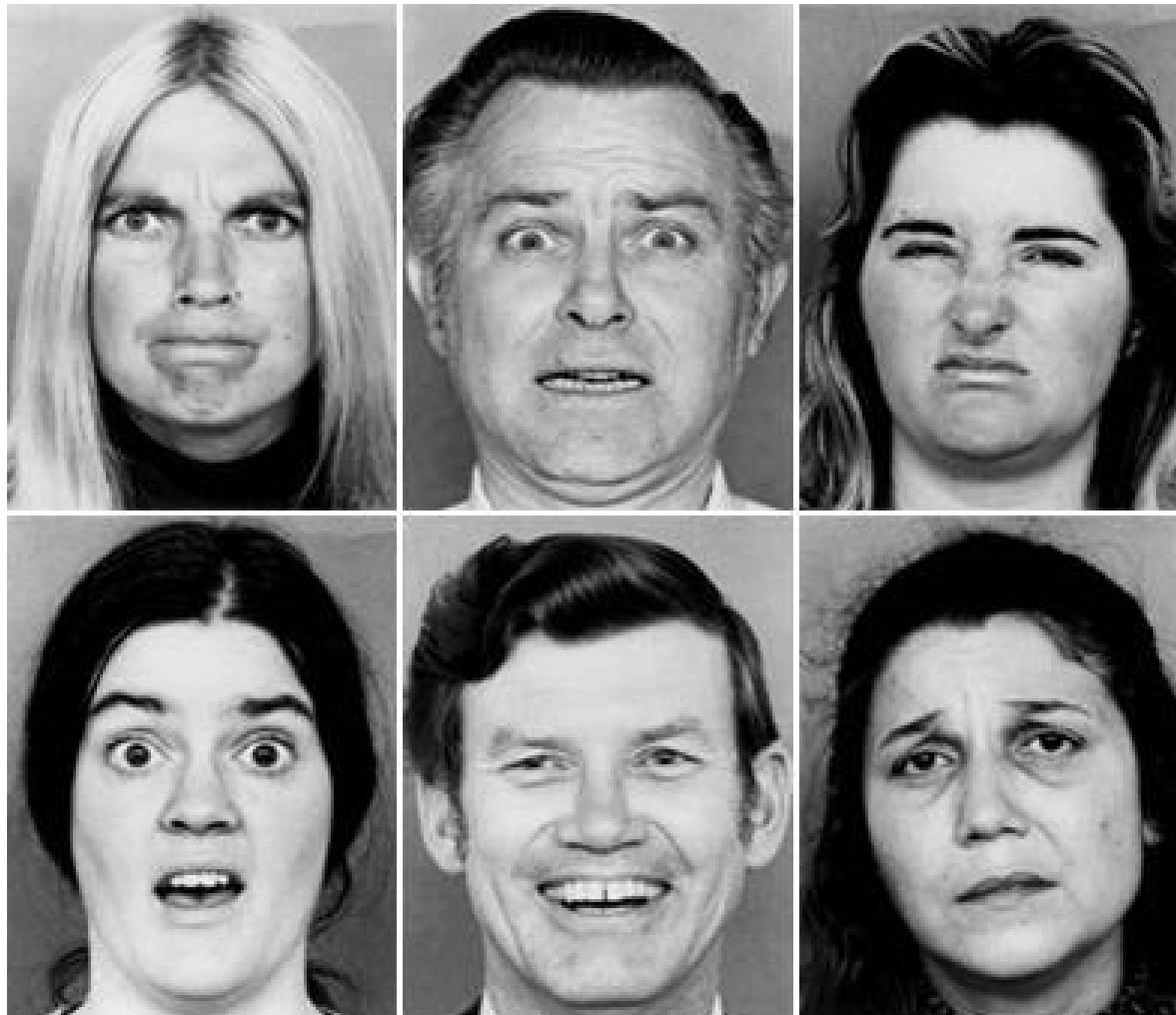
- Darwinian (e.g., Ekman, Izard, Frijda)
 - adaptive mechanisms, phylogenesis; basic set; expressions
- Jamesian (e.g., Levenson)
 - physiological changes -> subjective experience
- Cognitive (e.g., Ortony)
 - conscious aspects, based on appraisals
- Social Constructivism (e.g., Averill, Lunt)
 - social products and roles, fully explained at social level
- Neurobiology (e.g., Damasio, LeDoux)
 - understand fundamental brain mechanisms & subsystems
 - Involvement of emotions in decision-making and social intelligence (Damasio)
 - Double pathway to process emotional stimuli (LeDoux)

Which emotions?

- **Discrete categories** (evolutionary approach):
 - Primary:
 - Automatic, universal, role in survival
 - Fear, anger, happiness, sadness, disgust, surprise (expectation, acceptance)
 - Secondary:
 - Combination of more simple emotions
 - Feelings and (acquired) connections between object/situation categories and primary emotions
- **Continuum, 2/3 dimensions:**
 - Positive/negative (valence)
 - Activation/tranquility (arousal)
 - Powerfulness/powerlessness (potency)

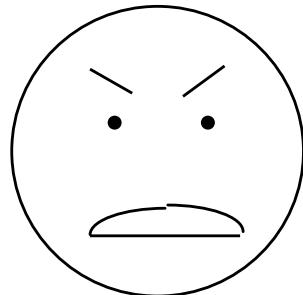
Discrete categories (basic emotions)

Anger, disgust, fear, joy, sadness, or surprise ?

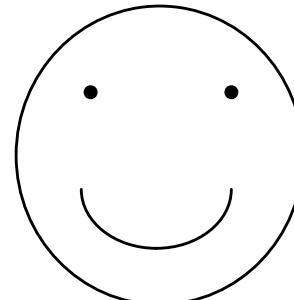


Discrete categories (basic emotions)

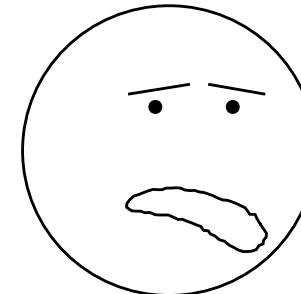
anger



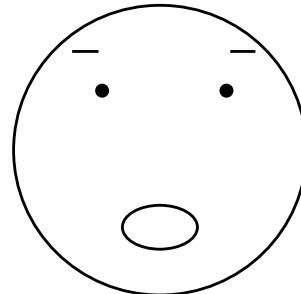
happiness



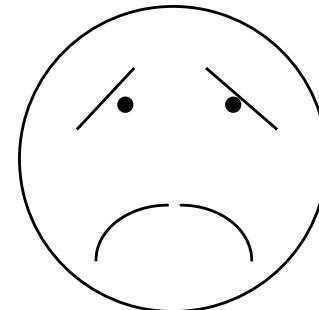
disgust



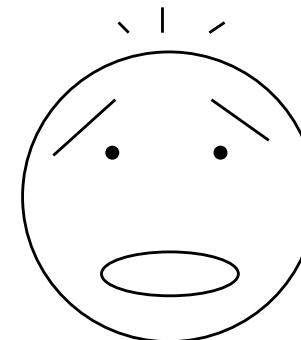
surprise / interest



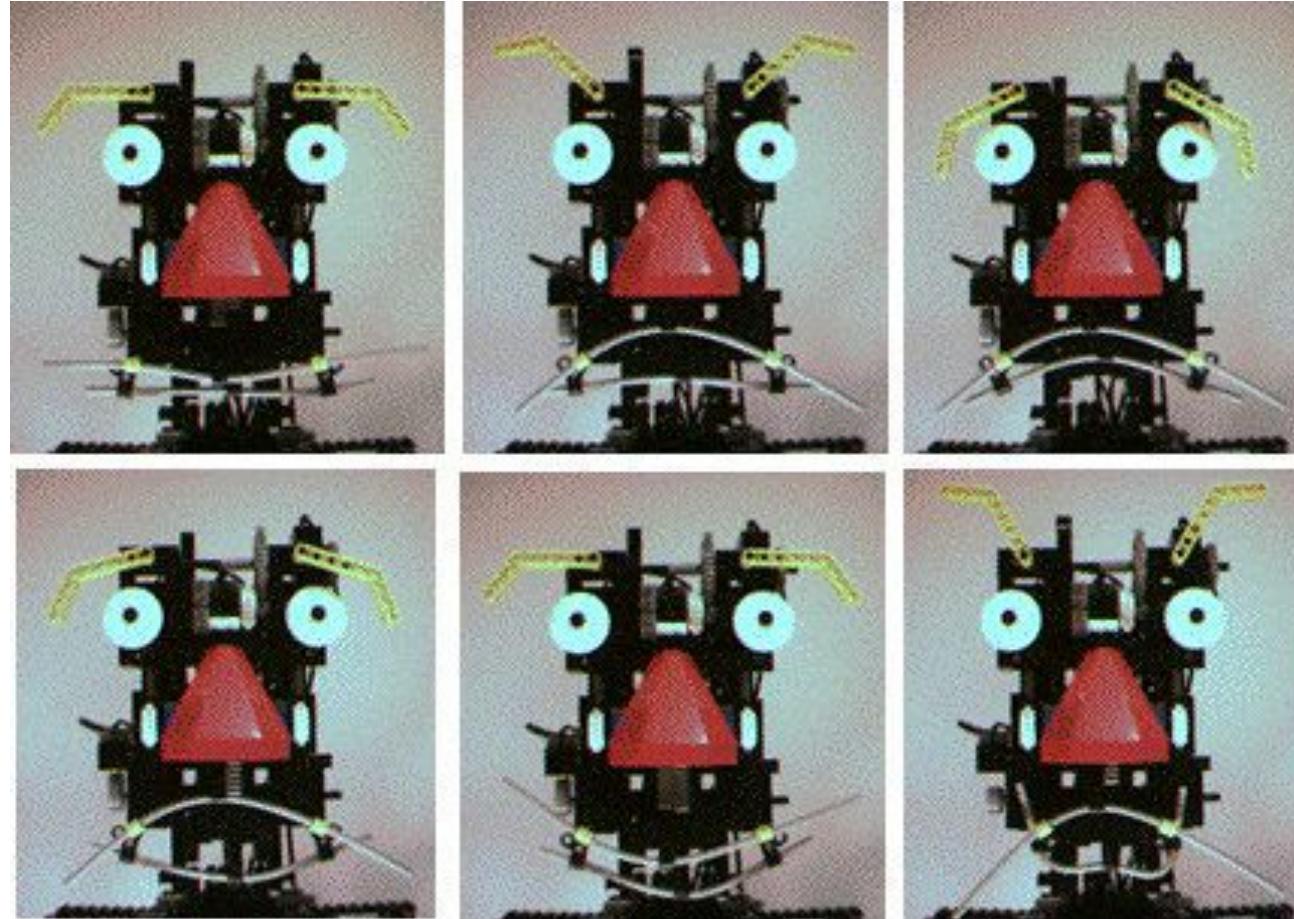
sadness



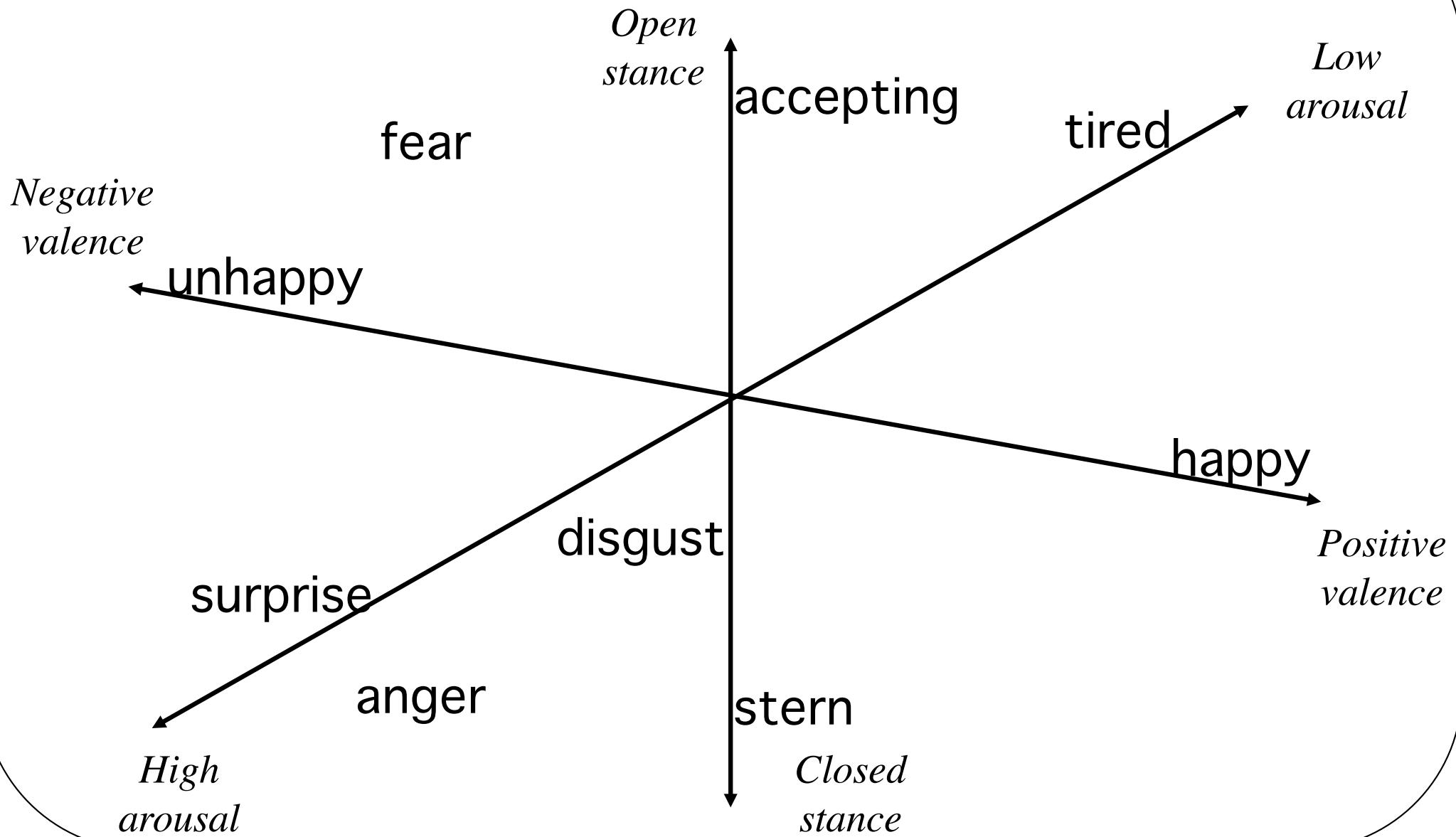
fear



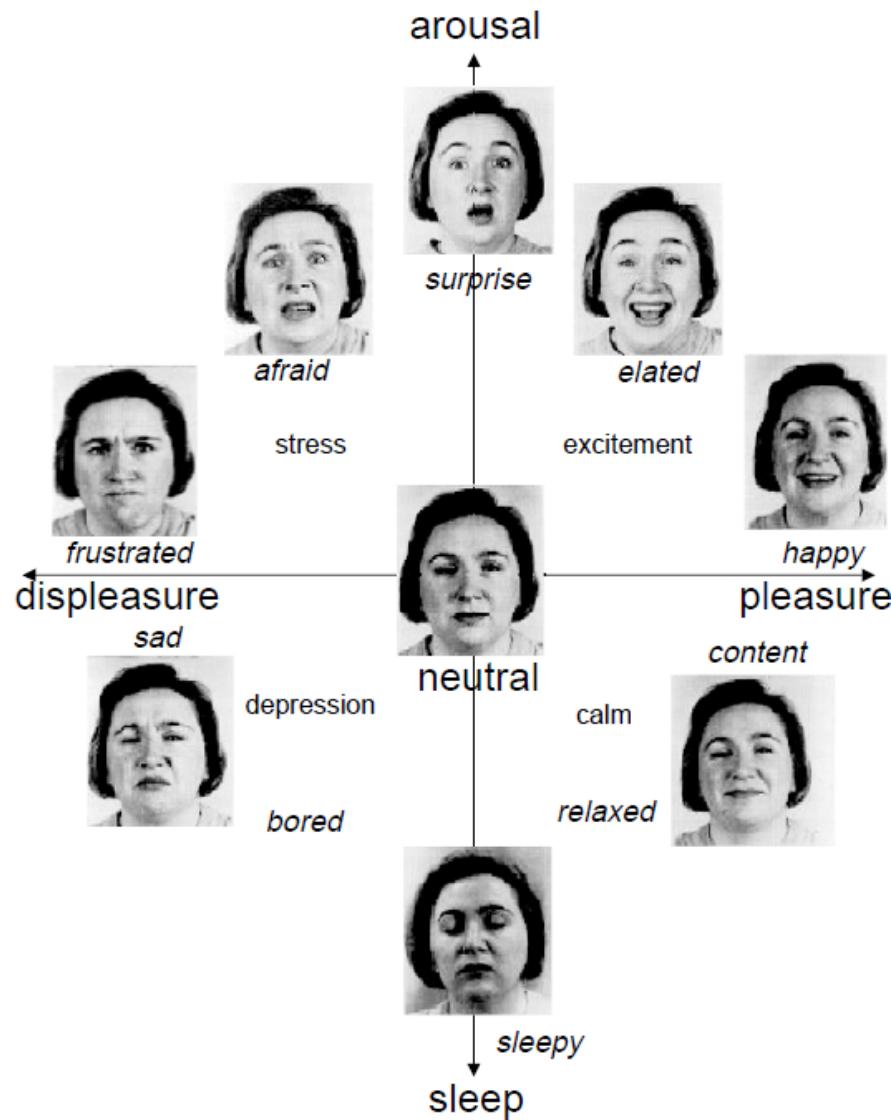
Feelix's emotional expressions



Dimensions of emotions

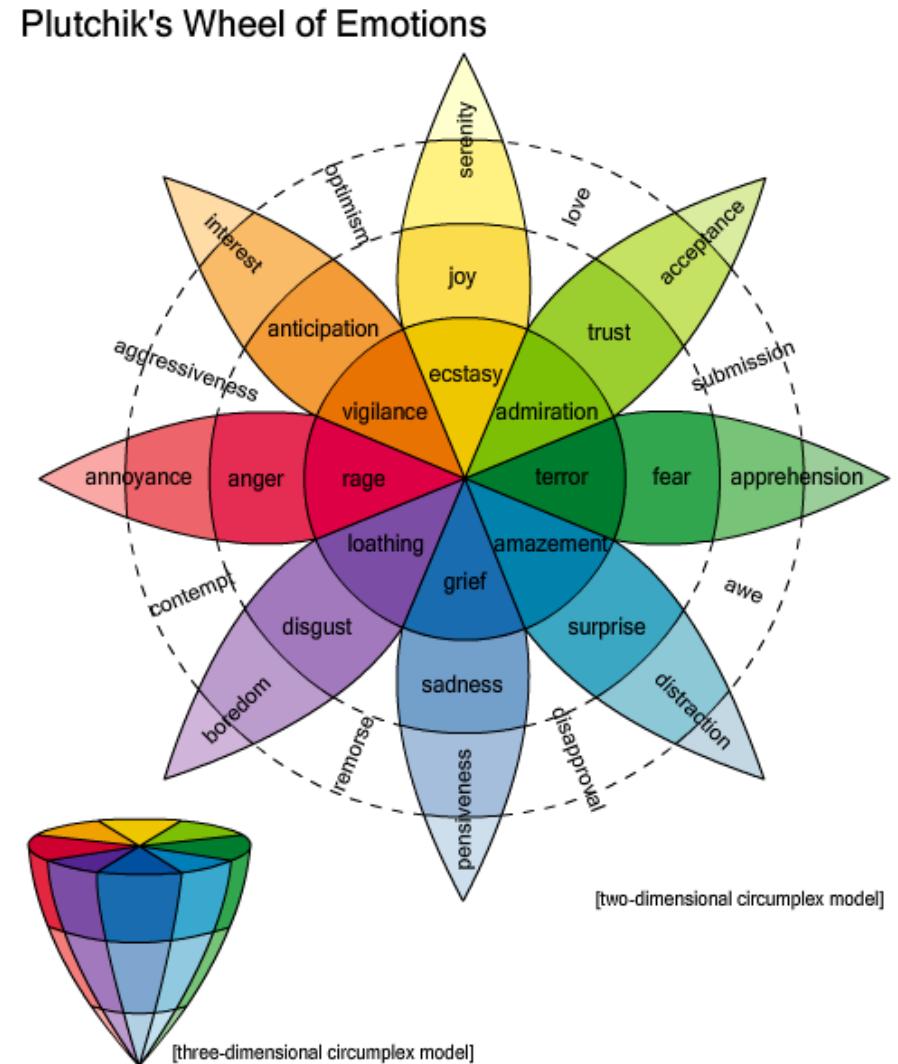


Dimensions of emotions



Plutchik's circumplex model

- 3D circumplex model of emotions
- Analogous to color wheel
- 8 Primary bipolar emotions
- Different intensities of emotions
- Combines discrete and dimensional models
- Can be mixed (secondary emotions)



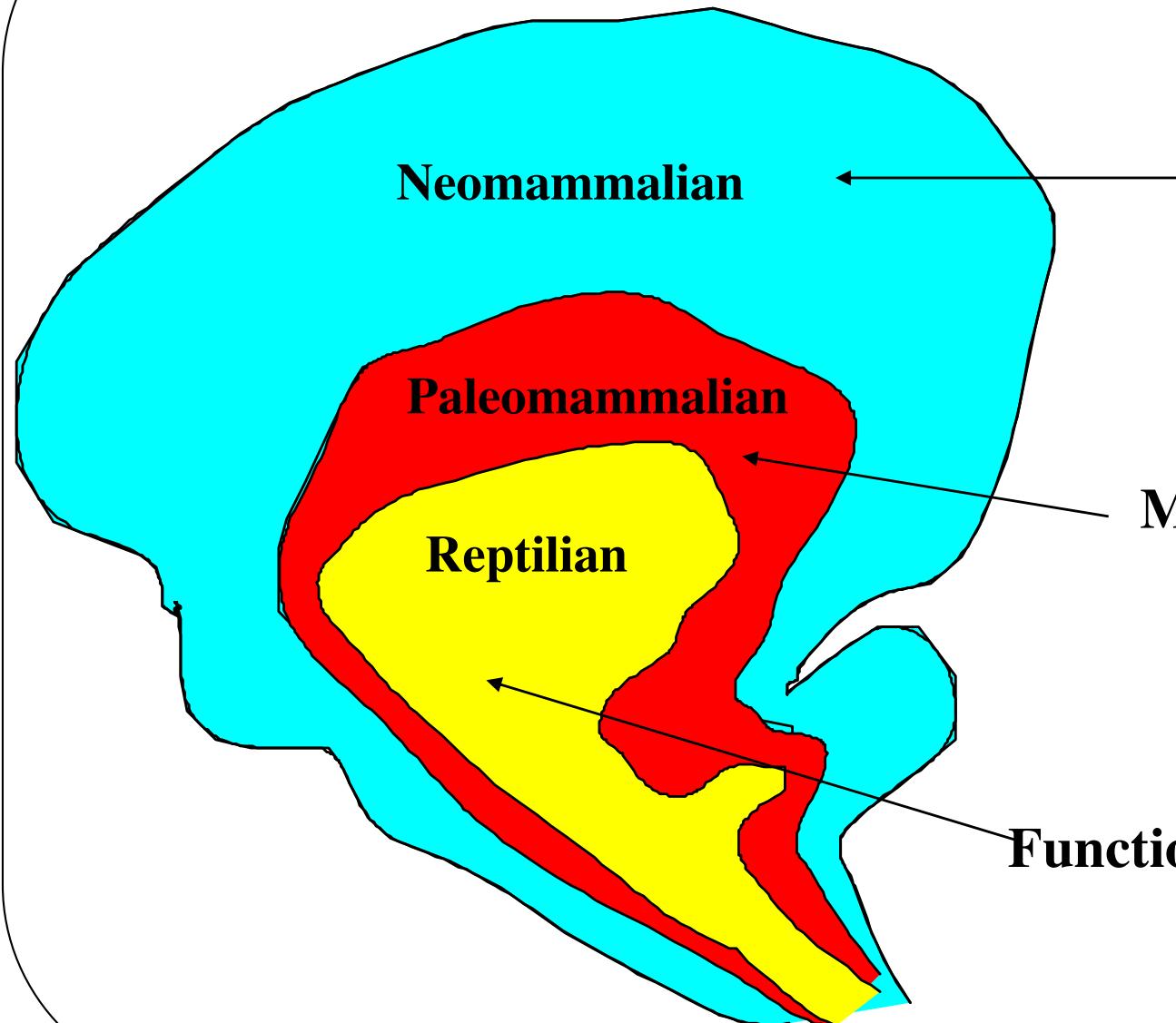
Why are emotions interesting for robots?

- **Adaptive mechanisms**
 - To deal with important events related to survival
 - Internal equilibrium, contribution to homeostasis
- **Motivate and guide action**
 - Categorize events as pleasant/unpleasant
 - Amplify/modify motivation
 - Decision making and social interactions (Damasio)
- **Value as signals**
 - Social reference (assess situation from others' emotions)
 - Communication (alarm calls: situation- and receptor-specific)
 - Construction of inter-subjectivity (image of self and other)

Characterizing emotions

- “Emotion integrates causally related processes from various subsystems. As in psychoanalysis, we see emotions as processes to stress their dynamic nature. The processes of the subsystems which are integrated in an emotional process are the following ones: *physiological, cognitive-evaluative, communicative-expressive, and subjective experience.*”
(Mandler 85, Pfeifer 91)
- “**Emotion** is the combination of a *mental evaluative process*, simple or complex, with *dispositional responses to that process*, mostly *toward the body proper*, resulting in an emotional body state, but also *toward the brain itself* (neurotransmitter nuclei in brain stem), resulting in additional mental changes. Note that, for the moment, I leave out of emotion the perception of all the changes that constitute the emotional response. ... I reserve the term **feeling** for the experience of those changes.”
(Damasio 94)

McLean's Triune Brain



**Planning, anticipation,
“cognitive functions”**

*Components: cortex and neocortex
(perceptual, motor, associative areas
prefrontal cortex)*

**Motivation, emotion,
attention, memory**

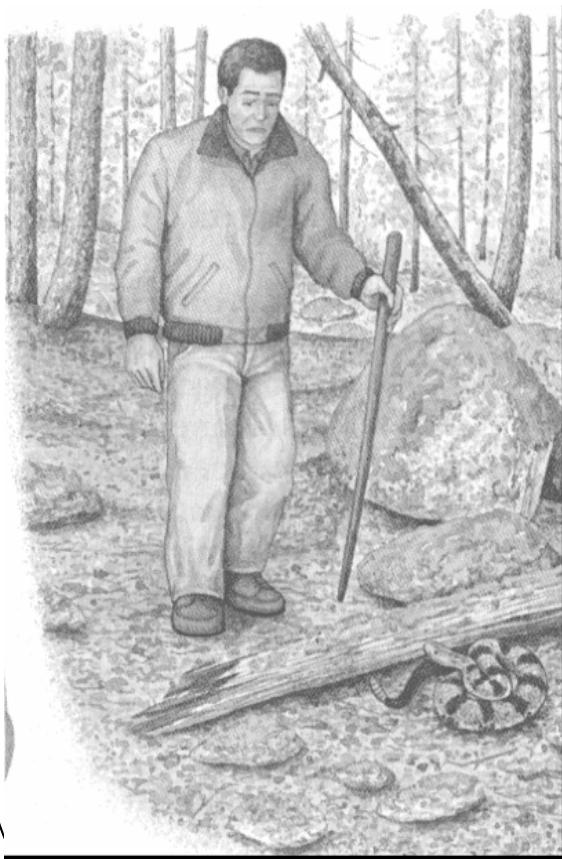
*Components: limbic system
(hypothalamus, amygdala, septum
cingulum, hippocampus etc.),*

**Functions related with regulation
and survival**

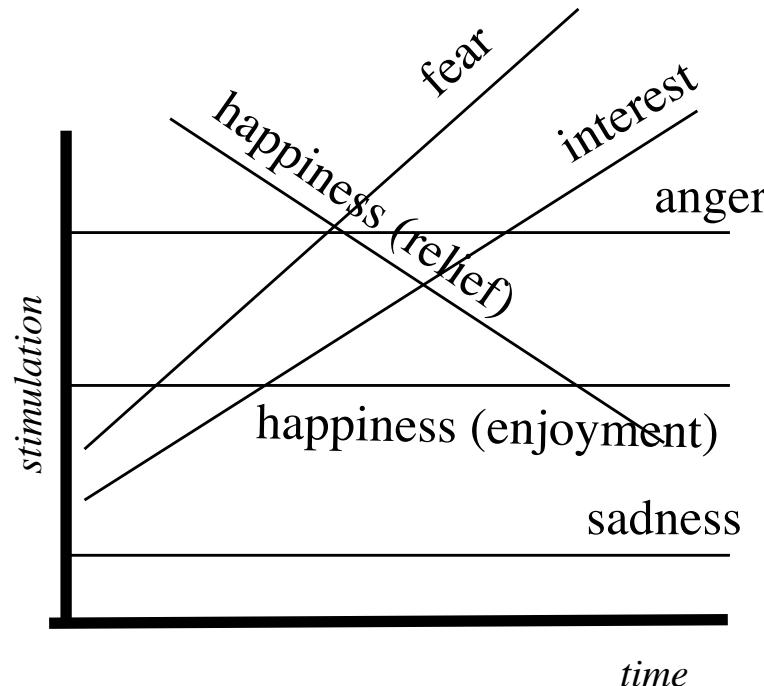
*Components: brain stem,
mesoencephalon, basal ganglia*

Emotion activation

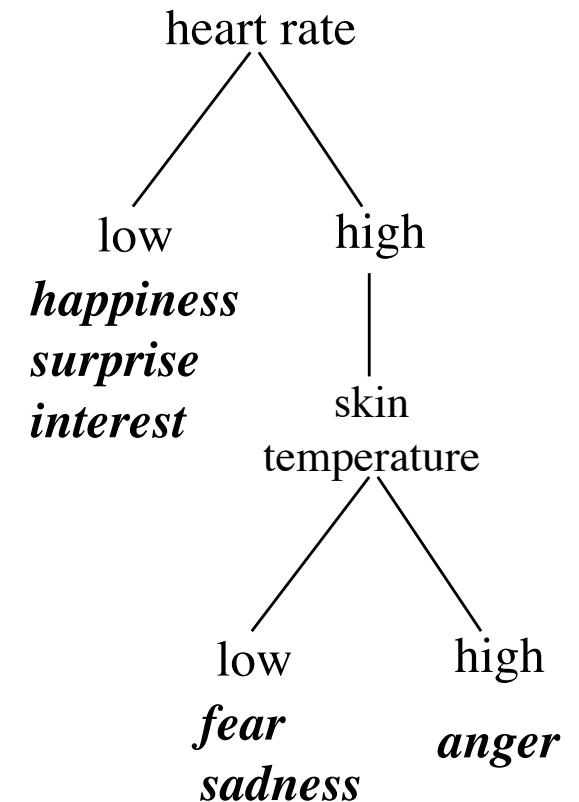
Significant event



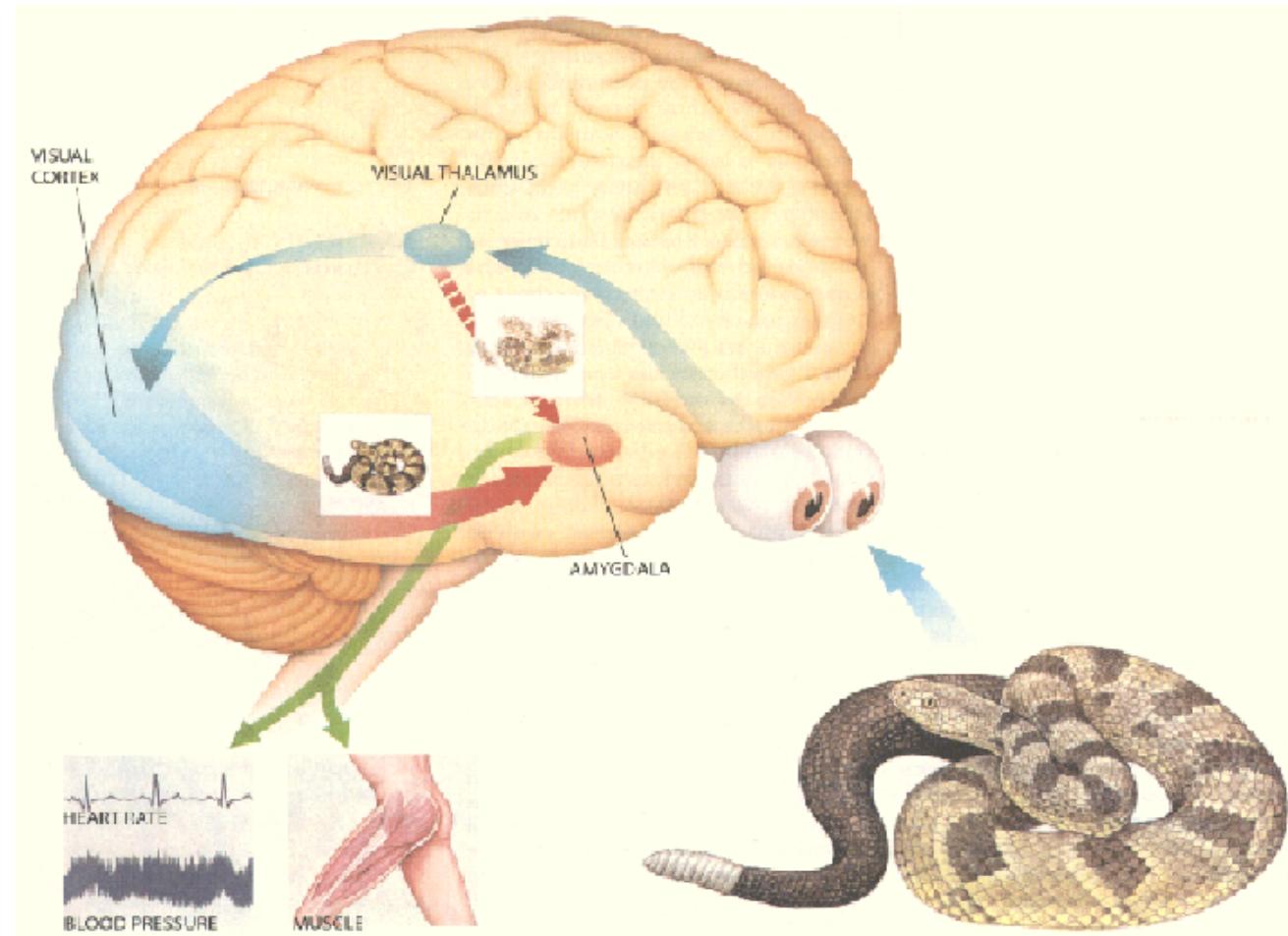
General patterns



Specific patterns



Example: fear



Emotions interest roboticists ...

- **Mechanisms for adaptation**
 - To deal with important events related to survival
 - Internal equilibrium, contribution to homeostasis
 - Control of interactions with (physical and social) environment
 - **Motivate and guide action**
 - Categorize events as pleasant/unpleasant
 - Amplify/modify motivation
 - Decision making and social interactions
 - **Value as signals**
 - Social reference (assess situation from others' emotions)
 - Interaction, communication
 - Construction of inter-subjectivity (image of self and other)
- “*Two-sided*” phenomena, connect “*inner world*” & “*external*” behavior

Agents, social agents, and emotions

- **Autonomous robots:**
 - Action selection
 - Navigation, exploration of environment
 - Reinforcement learning
- **Social robots:**
 - Collective and emergent behaviors
 - Service, human assistants (museum guides, etc.)
 - Cognitive mechanisms for social intelligence: learning by imitation, joint attention, etc.
- **Emotional robots:**
 - Single robot:
 - Architectures for behavior control (action selection)
 - Emotions in learning
 - Robot interacting with humans:
 - Emotion expression and recognition
 - Attachment bonds, empathy

Autonomous robots: common problems

- **Management of goals/motivations:**
 - How to change motivations' priorities?
 - “Weighting” motivations; emergency situations
- **“Loops” – repetitive and inefficient behavior:**
 - How to detect and avoid them?
- **Management of autonomy and learning:**
 - Deciding what (not) to learn
 - How to guide learning?
- **Memory (recall):**
 - Too global: cognitive overload, too long recall times
 - Need for selective recall (e.g., mood-congruent)

Emotion architectures for robots?

- “*What emotions are about is action (or motivation for action) and action control*” (Frijda, 1995)
- Emotions as **mechanisms to modify/maintain the relationships** between an **agent** and its **environment**:
 - Blocking influences from environment (*anger*)
 - Protection against these influences (*fear*)
 - Stop active relation (*sadness*)
 - Diminish risks of dealing with unknown environment (*anxiety*)
 - etc
- Emotions as a **2nd-order “monitoring” system** (internal and external environment)

Controlling (inter)action: examples

- **Happiness:**
 - Reinforcement of social bonds: produce openness towards others
 - Its expression indicates a disposition towards friendly interaction
 - Antidote against stress, “tonic”
- **Sadness:**
 - Reinforce social bonds: produce emphatic reactions, “ask for help”
 - Delay activity in motor and cognitive systems (change perspective, plan)
 - Signal that a problem is present
- **Anger:**
 - Deal with potential
 - Mobilize energy
 - Direct bodily resources towards action muscles
 - Its expression can prevent aggression
- **Fear:**
 - Motivate escaping from dangerous situation
 - Its anticipation induces avoidance behavior

Types of embodied emotion architectures

- Epiphenomenal (“emergent”) emotions
- Emotion “centers” / “subsystems”
- Emotions as dynamical systems
- Emotions as “functioning modes” via (“hormonal”) modulation of underlying neural circuitry

Emergent: Braitenberg's Vehicle 2

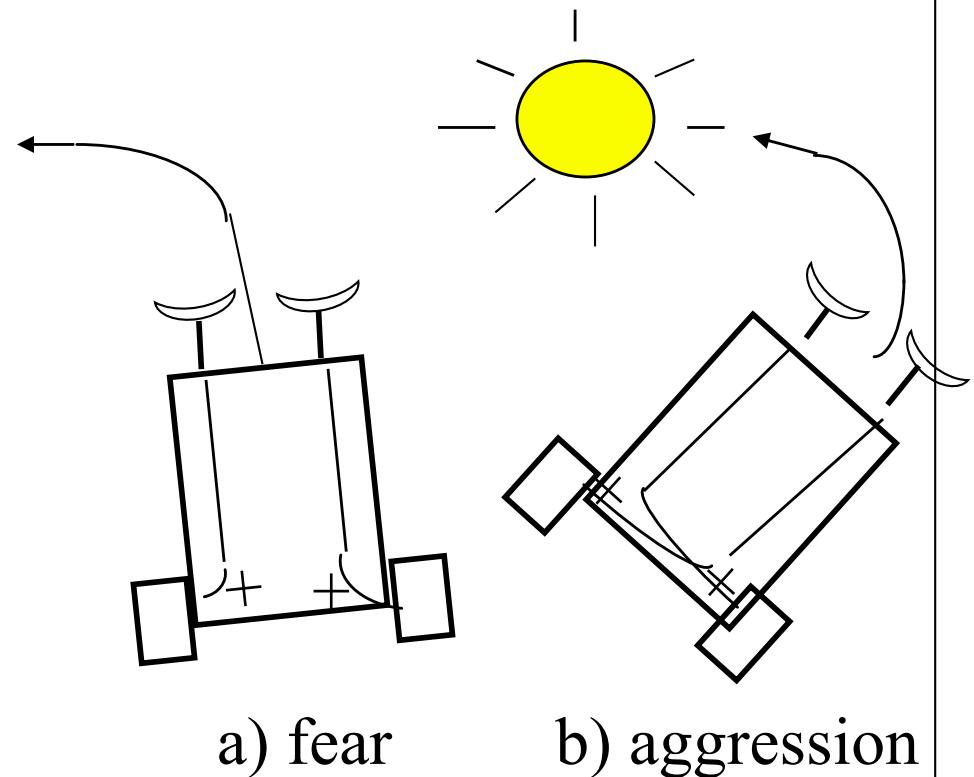
Machinery:

2 sensors, 2 motors
Connections:

- lateral (v2a)
- counter-lateral (v2b)

Behavior:

- Frontal source: hit source
- Source laterally located:
 - a) Away from source (positive feedback.)
 - b) Approach source (negative feedback)



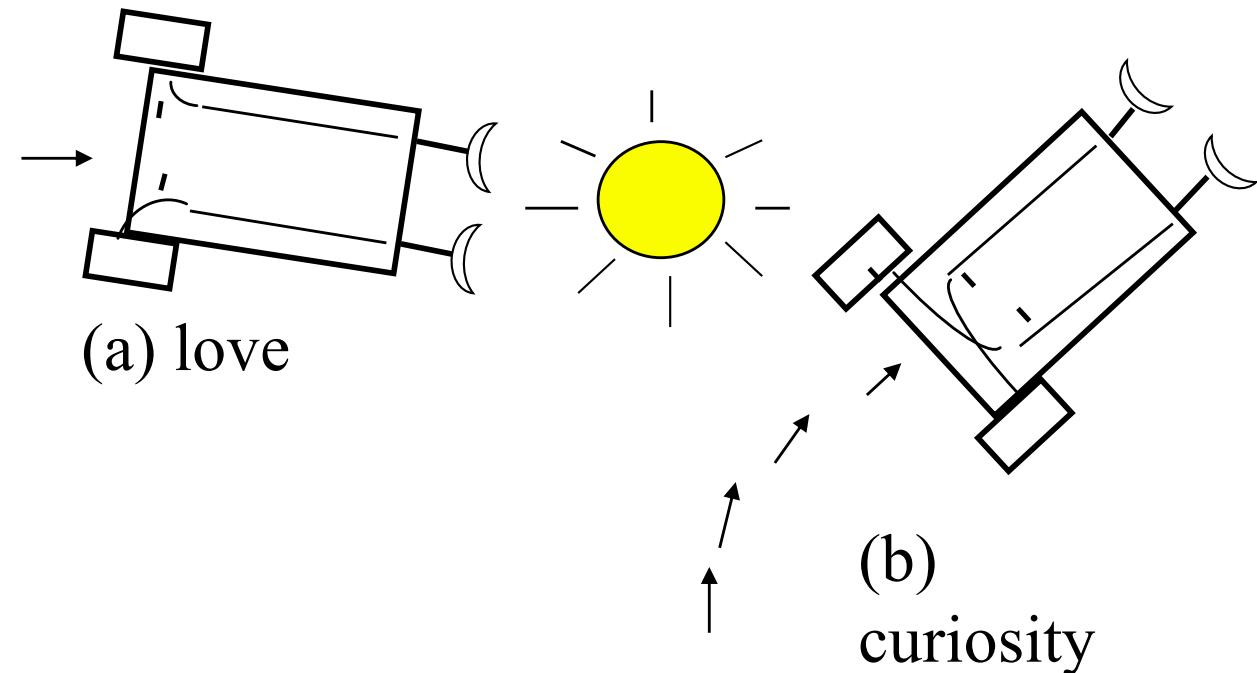
Emergent: Braitenberg's Vehicle 3

Machinery:

Sensors inhibit motors

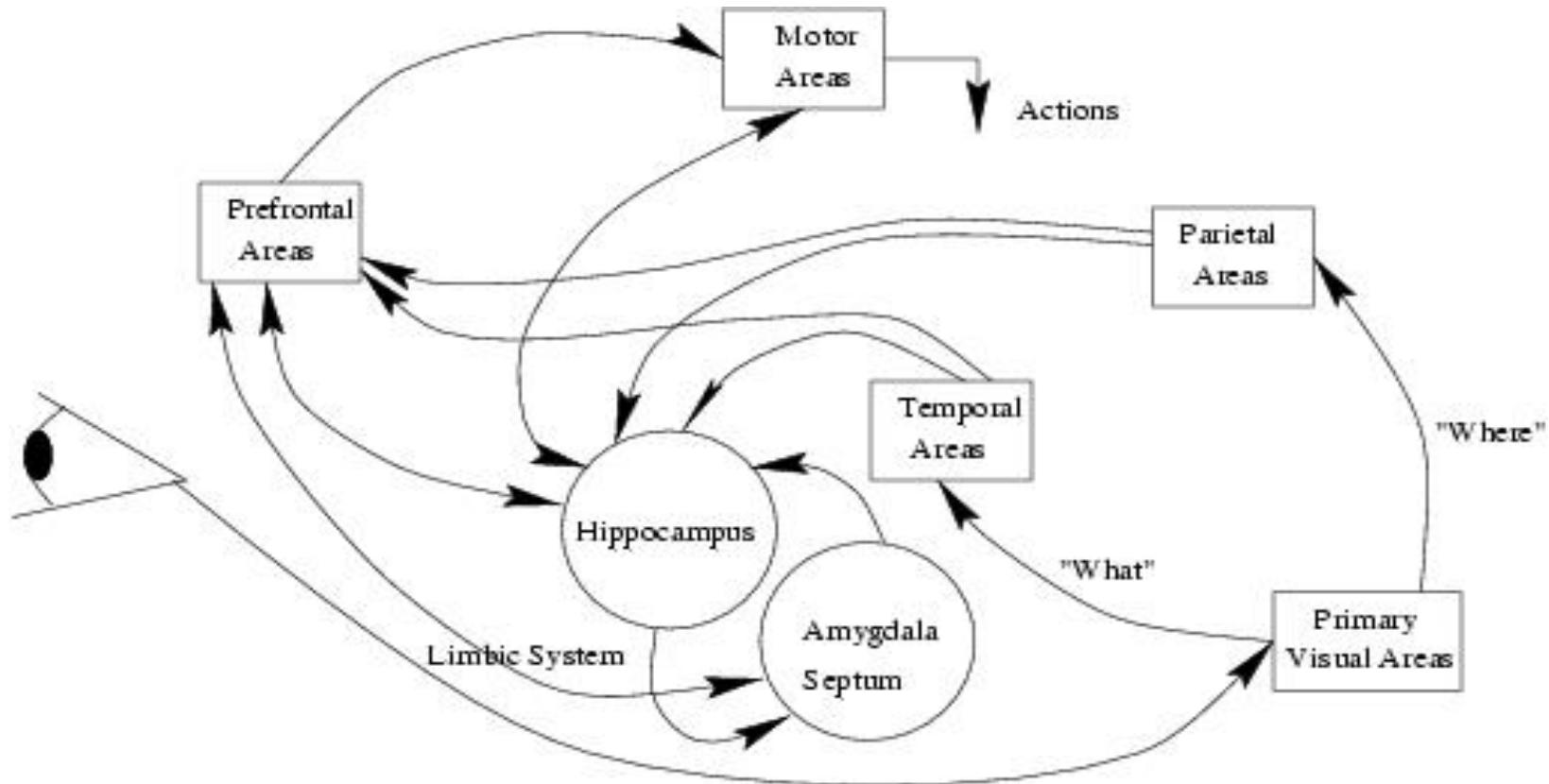
Connections:

- lateral (v3a)
- counter-lateral (v3b)



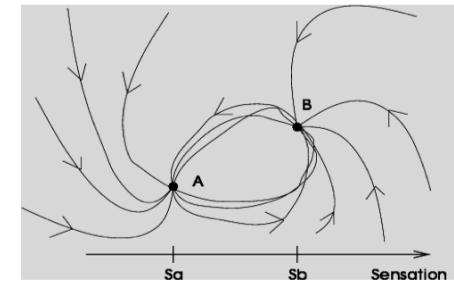
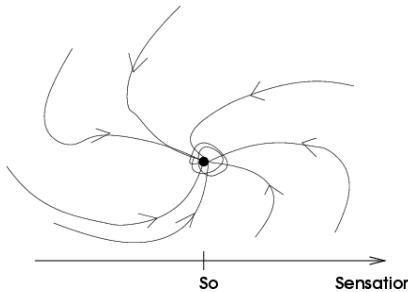
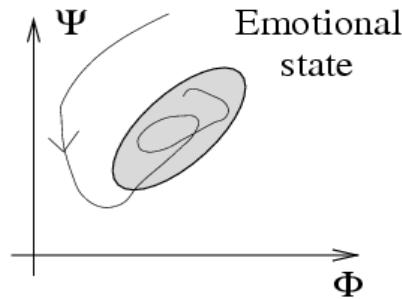
- ◆ Emotions “in the eye of the beholder” approach:
 - ◆ Prevents over-design and over-attribution
 - ◆ Fails to account for how biological emotions relate and influence different cognitive and behavioral subsystems simultaneously

Emotion “centers” / “circuits”



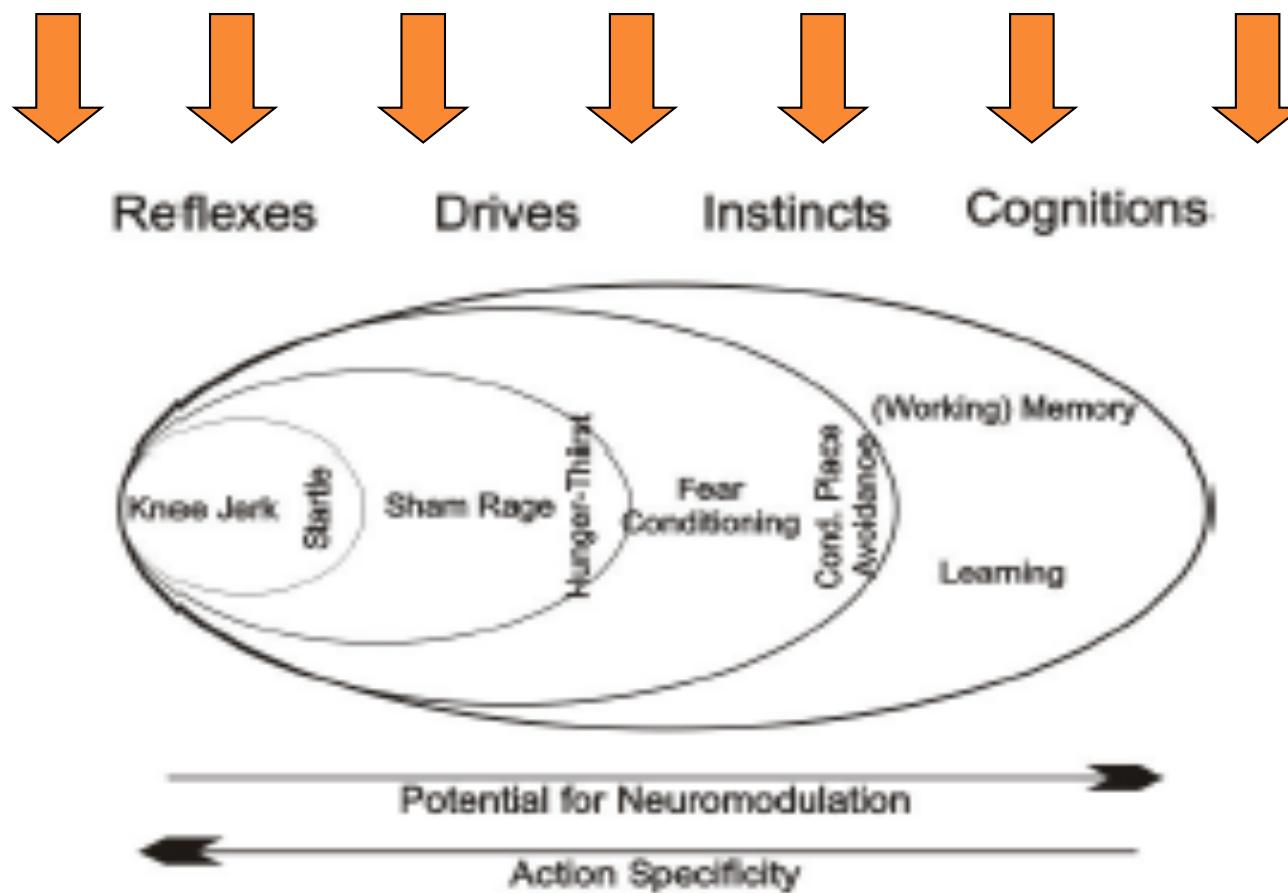
- How does the artificial Nnet “match” a natural one?
- What is modeled?
- How to relate functions to underlying structure?
- How to relate functions?

Emotions as dynamical systems



- Oscillators, attractors, etc to model “internal” and “interactive” aspects of emotions
- DS to bridge gap between (among) disciplines, Marc D. Lewis, “Bridging emotion theory and neurobiology through dynamical systems modeling”, *Behavioral and Brain Sciences* (2005) 28, 169-245.
- Some issues:
 - Everything explained as a DS ???!
 - Different disciplines, same vocabulary, ... same meanings?
 - DS as “metaphor” or mathematics?

Emotions as patterns of neuromodulations

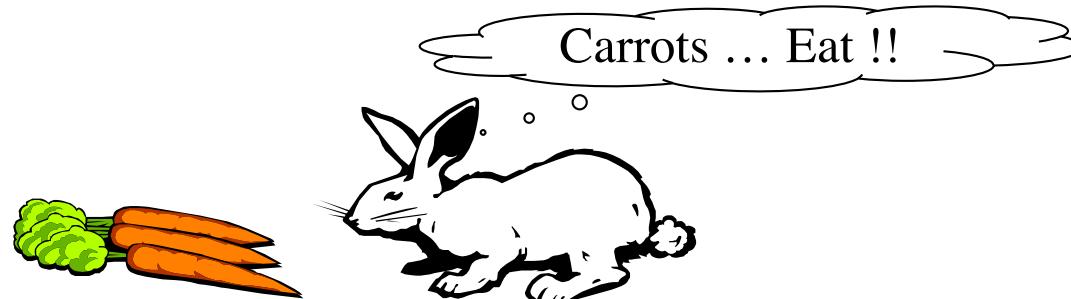


(after Jean-Marc Fellous)

- Emotions: different “functioning modes” of the nervous system
- Emotions can easily influence cognition and action
- Increasing potential for neuromodulation: incremental study

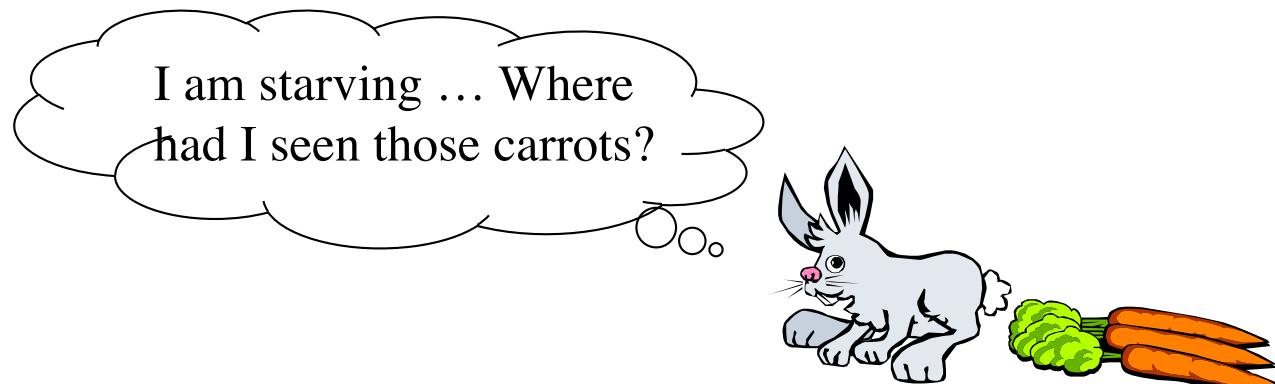
Animals “want” & “like”: decision making

1 ...



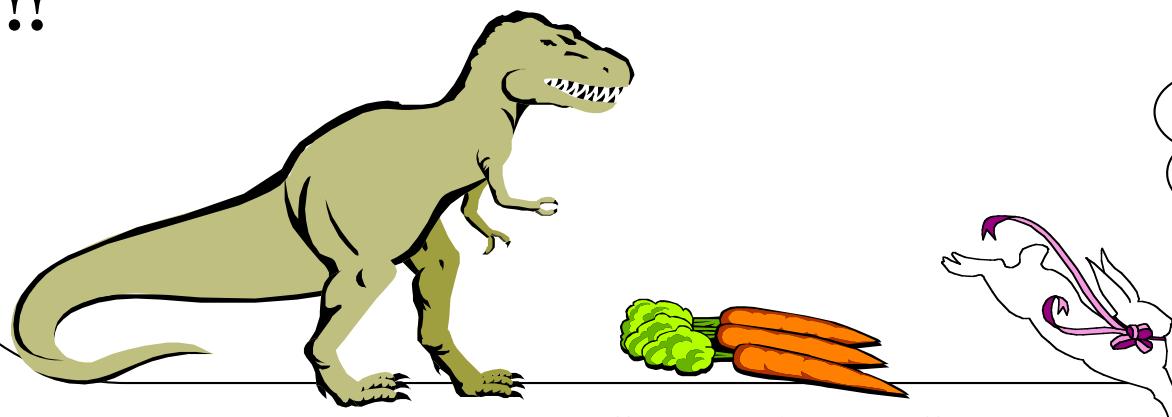
reactive system

2 ...



motivated system

3 !!



emotional system

Bibliography

- Cañamero, L.D. 1997. Modeling Motivations and Emotions as a Basis for Intelligent Behavior. In W. Lewis Johnson, ed., Proceedings of the First International Conference on Autonomous Agents, 148–155. New York: ACM Press
- Cañamero, L.D. 2003. Designing Emotions for Activity Selection in Autonomous Agents. In R. Trappi, P. Petta, S. Payr, eds., *Emotions in Humans and Artifacts*. Cambridge, MA: MIT Press.
- Cañamero, L. and Gaussier, P. 2005. Emotion Understanding: Robots As Tools and Models. In J. Nadel and D. Muir (Eds.), *Emotional Development: Recent research advances*. Oxford University Press.
- Picard, R.W. 1995. Affective Computing, MIT TR-321.
- Picard, R.W. 1997. *Affective Computing*. Cambridge, MA: The MIT Press.