Possible Solutions from Robotic Models of Emotion

(Identifying theoretical directions to resolve problems)

Lola Cañamero
Adaptive Systems Research Group
University of Hertfordshire, UK

L.Canamero@herts.ac.uk
Problems?

• Models
  – Scope of different theories / models? Compatible?
  – What is meant by “emotion” in each of them?
  – *theoretical notions too abstract, conceptual confusion, need to operationalize*

• Emotion “machinery”
  – Plausible mechanisms underlying different aspects of emotions and their influence in cognition and action?
  – How much “emotion machinery” do we need to postulate (and for what)?
  – *need for a synthetic and incremental approach*
Problems? (2)

• Applications
  – Which emotions / aspects for different applications?
  – What guidance can emotion theories provide?
  – > need for collaborative work

• Assessment of the influence of emotions
  – How can this be measured (a) from “inside” the organism/architecture; (b) from observed behavior?
  – To what extent does “emotional behavior” respond to specific internal “emotion machinery” / can be explained as a side-effect of the interactions?
  – > need to develop precise metrics, standards
Why robots?

- **Synthetic approach**
  - Models must be precise to build systems (robots as models)
  - Robots as tools, “laboratories” for systematic testing and experimentation
    - Controllable, manipulable parameters
    - Normal and abnormal behavior
  - > *new challenges and feedback to emotion theorists*

- **Physical devices**
  - Interact autonomously in the physical (“real”) world
  - Uncertainty, dynamism, noise, dynamics of interaction -> side-effects
AI and Cognition/Action

- Sensory motor level was considered as secondary
  - Important debate in 1980/1990
- Embodied AI approach
  - Situated cognition and action
  - Sensory-motor dynamics
  - Motivation
  - Autonomy
  - ...
- Tendency to assign intentions to others / artifacts

*Do emotions “exist”? Are they useful?*
Building a cognitive map

Controlled by internal motivations
No need of emotions?
Gesture learning
Why emotions interest roboticists…

• Adaptive mechanisms
  – To deal with important events related to survival (LeDoux)
  – 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 (Damasio)

• Value as signals
  – Social reference (assess situation from others’ emotions)
  – Communication
  – Construction of inter-subjectivity (image of self and other)

➢ Dual aspect, connecting “inner world” and “external” behavior
Some problems in robotics…

- **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)

-> *emotions could improve all these problems; how to proceed?*
1. Problems regarding models

– Scope of different theories / models? Compatible?
– What is meant by “emotion” in each of them?
→ *theoretical notions too abstract, conceptual confusion, need to operationalize*
Example: behavior selection

1 ...

Carrots … Eat !!

reactive system

2 ...

I am starving … Where had I seen those carrots?

motivated system

3 !!

I am starving, but this might not be the right moment!!!

emotional system
Example: functional / motivational model

- *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)
Motivation+Emotion architecture

Perception

- stimuli
- environment

Control

- internal milieu
  - controlled variables, “hormones”
- motivations
  - (hunger, thirst, fatigue, protection, ...)

Action

- behaviors
  - (walk, withdraw, eat, drink, attack, ... go-toward, look-for ...)
- activity

Emotions

- (fear, anger, happiness, sadness, boredom, interest)
<table>
<thead>
<tr>
<th>Problems with reactive and motivated architectures</th>
<th>Contributions of emotions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Rigidity of behavior (S → R)</td>
<td>• More flexible and varied behavior as a function of internal state (e.g. predator: attack or flee)</td>
</tr>
<tr>
<td>• Insufficient autonomy (reactions, drives)</td>
<td>• Modulation/change of motivational state and behavior</td>
</tr>
<tr>
<td>• Repetitive and inefficient behavior (loops)</td>
<td>• Self-monitoring, interruption of inefficient behavior</td>
</tr>
<tr>
<td>• Inefficient treatment of urgency situations, “goal forgetfulness”</td>
<td>• Faster responses, anticipation (emotional memory)</td>
</tr>
<tr>
<td>• “Atomic” behavior</td>
<td>• Re-equilibration of internal milieu, back to goal</td>
</tr>
<tr>
<td></td>
<td>• Behavioral chains (e.g. fear, escape, anger, attack, relief)</td>
</tr>
</tbody>
</table>
Issues raised

• Different theories combined:
  – relations between them?
  – Which other theories would be “compatible” (e.g., emotion elicitation, appraisal?)

• Categories / dimensions?

• Valenced reactions:
  – Where does valence come from?
  – Grounding in robots? Pleasure / pain?
The example of imitation
Bootstrap for low-level imitation

- Proto-imitation as a side effect of a homeostatic system (perceptual ambiguity)
- No need of a special imitation “module”
Imitation and social interactions (2)

Double function of imitation:
- Learning
- Communication

Capability to predict the rhythm of the exchange can be used as an internal reward (“pleasure”)

Interest of displaying the internal state induced by reward (second order interaction) How to express it quickly?
2. Problems regarding “emotion machinery”

– Plausible mechanisms underlying different aspects of emotions and their influence in cognition and action?
– How much “emotion machinery” do we need to postulate (and for what)?
→ need for a synthetic and incremental approach
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)

“Moral”:

Intentional terms used to describe the vehicle’s behavior are in the eye of the beholder
Braitenberg’s Vehicle 3

Machinery:

Sensors inhibit motors
Connections:
- lateral (v3a)
- counter-lateral (v3b)

Behavior:
- a) Orientation towards source, slows down until stops in front of it (negative feedback)
- b) Slows down towards source, turns away when close, environment can become more attractive

“Moral”:

Intelligence emerges from the interactions of the vehicle with its environment and from the interactions among the creature’s components
Plausible mechanisms?

• “Emergent” emotions (observer)
  – Prevents over-design and over-attributeation
  – Misses the ability of emotions to relate and influence
different cognitive and behavioral subsystems
simultaneously

• Emotion “centers”
  – How to relate functions to underlying structure?
  – How to relate functions?

• Neuromodulation of “neural circuitry” / cognition &
action
  – Emotions can easily influence cognition and action
  – Increasing potential for neuromodulation: incremental study
  – However, what produces “neuromodulations”? -> appraisal?
Increasing complexity...

1. A 7-neuron version of the type 2 Braitenberg vehicle that changes from light-avoiding to light-seeking behaviour by modulation of its PDP interneuron transfer functions.

2. A variant on 1 that includes the ability to signal to other vehicles the onset of a change in its behaviour. This vehicle also includes the ability to receive and so be affected by the same signal broadcast by a conspecific.

3. A variant on 2 that includes the ability to sensitise the vehicle’s sensory system and so make it more sensitive to stimuli from its environment.

4. A variant on 3 that uses the notion of ‘Primers’ and ‘Releasers’ to choose between a larger range of competing behaviours in a ‘behaviour system’.

5. A variant on 4 that uses the notion of a ‘Value system’ to learn useful mixes of modulators to affect its behaviour system.

6. A variant on 5 in which the input to the value system is also affected by modulators.

7. In contrast to the above ideas, here we start to use evolution to reduce the amount of human input (and so prejudice) in designing a vehicle.

8. A variant on 7 that explores neuromodulation as the basis of ontogenetic development for a ‘young’ vehicle.
3. Problems regarding applications

– Which emotions / aspects for different applications?
– What guidance can emotion theories provide?
–> need for collaborative work
Choice of model: categories?

anger  happiness  disgust

surprise / interest  sadness  fear
Complexity: caricature or portrait?

Emot. reaction, familiarity

+ 100%

Similarity to human

0

moving corpse

100%

“healthy” people

toy robot

uncanny valley
Feelix, a simple “humanoid”
Feelix’s basic expressions
Feelix’s emotion model

• **Facial expressions:**
  – Division of face in 2 halves (upper, lower), **perceptual primitives**
  – Expressive **features** (4 DoF): eyebrows, mouth (2 lips)

• **Emotion activation through interaction:**
  – **Tactile** stimulation: duration and frequency of presses
  – General activation **patterns:**
    • Sudden stimulation increase: *surprise, fear*
      – few short and frequent presses
    • Sustained high stimulation level (over-stimulation): *anger*
      – many short and frequent presses / very long presses
    • Moderate stimulation: *happiness* (enjoyment)
      – “nice” presses (not too long, not too short, not too many)
    • Sustained low stimulation level: *sadness*
      – absence of / very few presses
Playing the emotion game with Feelix

• **Tests on recognition of emotional expressions**
  - Three tests: (1) free; (2) multiple-choice; (3) human faces
  - Results congruent with psychological studies

• **Interesting observations!**
  - *During emotion recognition:*
    • People mimicking Feelix’s face to identify the emotion
    • ”Empathy”: people mirroring the expression with face and vocal inflection while commenting (”Ooh, poor you”, ”Look, now it’s happy!”)
  - *During interaction with Feelix:*
    • People mirroring the emotion they wanted to elicit:
      - With their faces
      - In the nature of the presses applied to the feet
Our new expressive head
Some expressions
Tool for autism research

- Study of imitation and evocation of emotional expressions by typical and autistic children (P. Brun)
- Interesting features:
  - Adjustable, controllable parameters (number of features, movement, speed, timing, etc)
  - Wide range of expressions (prototypes & atypical)
  - Still / dynamic expressions
  - Display only or interaction
  - Similarity / dissimilarity to human?
    - Is a “humanoid” head perceived as human?
  - 3D, physical device (vs video, virtual reality, etc)
Choice of model: dimensions?

- **Choice of model:** dimensions?
- **Negative valence:**
  - fear
  - unhappy
  - surprise
- **Positive valence:**
  - happy
  - tired

- **High arousal:**
  - anger
  - stern
- **Low arousal:**
  - accepting
  - closed stance

**Open stance**

**Closed stance**
Kismet, a sophisticated head
4. Assessment of the influence of emotions

– How can this be measured (a) from “inside” the organism/architecture; (b) from observed behavior?
– To what extent does “emotional behavior” respond to specific internal “emotion machinery” / can be explained as a side-effect of the interactions?

--> need to develop precise metrics, standards
Internal: viability

Ashby’s Viability zone (Physiological space)

Set point

Critical zone

Ideal zone

Lethal boundary
Viability indicators

- **Life Span**: The time that the robot survived (remained viable) during each run in simulation steps normalized with the total simulation time.

- **Overall Comfort**: The average level of satisfaction of all the essential variables normalized between [0,1].

- **Physiological Balance**: The homogeneity with which the different physiological needs are satisfied.

- **Management of death risk**: The amount of time during which an architecture keeps the highest physiological need within different regions of the viability zone.
Activity cycles

- Over-Opportunism
- Goal Interference
External: Behavior analysis

- Ethological methods pick up the manifestations of “side-effects” in a systematic way.
- Some of these “side-effects” are easily over-interpreted by human observers as reflecting specific “drives”, “motivations” or “emotions”.
- Can help to prevent over-attribution.
- To what extent does the analysis of observed behavior help us understand …
  - underlying emotional mechanisms?
  - features of the observer’s perception (of emotions)?
Sparky, a tele-operated robot
Conclusion

• Build models beyond the level of phenomena
  – Take into account emergent properties (structure ≠ function)
• Developmental approach
  – Sensory-motor learning
  – emergent property of a single architecture
• Emotions as second-order monitoring systems ; dual aspect

⇒ Two sides of a coin:
  • Robot as model to understand emotional / cognitive mechanisms and their deficits
  • Robot as tool for psychologists to test the validity of the model