Motivation and Affect in MicroPsi

MAS S66
New Destinations in Artificial Intelligence
Goals and Directions for Future Research

joscha@mit.edu
Components for Cognitive AI

- Memory
- Resolution
- Perception
- Action selection
- Action
- Securing rate
- Arousal
- Motivational system: Motive Selection and Decision making
- Selection threshold
- Urges/drives
Layers of Cognition

- Reflective
- Deliberative
- Reactive
Columns of Cognition

- Perception
- Cognitive Processing
- Action
## Cognitive Grid

<table>
<thead>
<tr>
<th>Reflexive Perception</th>
<th>Meta-Management</th>
<th>Management Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deliberative Perception</td>
<td>Planning, Reasoning</td>
<td>Deliberative Action</td>
</tr>
<tr>
<td>Reactive Perception</td>
<td>Reflexes</td>
<td>Reflexive Action</td>
</tr>
</tbody>
</table>
Conceptual Analysis: HCogAff (Sloman 2001)

- Perception
- Meta-management (reflective processes)
- Deliberative processes (Planning, decision making, ‘what-if’ reasoning)
- Reactive processes
- Action
- Personae (self model)
- Long term memory
- Motive activation
- Alarms: real-time intervention

12/2/15 FutureAI Environment
Cognitive Artificial Intelligence

Methods should focus on components and performances necessary for intelligence:

- **Universal Representations:**
  Grounded neuro-symbolic representations (integrate both symbolic and distributed aspects)

- **(Semi-) Universal Problem Solving:**
  Learning, Planning, Reasoning, Analogies, Action Control, Reflection ...

- **Universal Motivation:**
  Polythematic, adaptive goal identification

- Emotion and affect

- Whole, testable architectures
Modeling Motivation in a Cognitive Architecture

- **General intelligence needs General Motivation**
- Motivational system structures cognition
- Motivational dynamics: physiological, social and cognitive drives
- Intention selection and action control
- Motivation vs. affect
MicroPsi architecture

PSI theory

Principles of Synthetic Intelligence
(Dörner 1999; Bach 2003, 2009)
Acknowledgements

Work on MicroPsi2 is collaborative effort:
- **Ronnie Vuine, Dominik Welland, Priska Herger, Jonas Kemper** are contributors to the current version
- Architecture/concepts have been inspired by Dietrich Dörner, Aaron Sloman, Marvin Minsky, Stan Franklin and many others
- Support from Humboldt University of Berlin, University of Osnabrück (Institute for Cognitive Science), Berlin School of Mind and Brain, Harvard Program of Evolutionary Dynamics, MIT Media Lab
MicroPsi Principles

- Neuro-Symbolic architecture
- Agents implemented as spreading activation networks
- Unified representations, different sets of operations
- All representations are grounded
- Meaning is attached to representations by motivation
Implementation: MicroPsi 2 (Bach, Welland, Vuine, Herger 12, 14)
Goals in MicroPsi

• Goal: situation or action that affords to satisfy a need
• Aversive goal: situation or action that frustrate a need
• All behavior is directed on satisfying an appetitive goal or avoiding an aversive goal
• Needs are predefined, goals are learned
From Needs to Behavior

Needs

Urge Signals

Priming + Modulation

Learning

Decision Making

Memory

Perception

Action
Motivation in MicroPsi

Pleasure and distress:

- Change of a demand is reflected in pleasure or distress signal
- Strength is proportional to amount of change

- Pleasure and distress signals deliver reinforcement values for behavioral procedures and episodic sequences and define appetitive and aversive goals.
Motivational System

• drive = demand + urge indicator

Water

$s_{water}$

urge indicator

current level

target
Motivational Learning

- motive = urge + goal situation

Diagram:
- Target
- Current level
- Water
- 只有 water
- Urge indicator
- Goal
• motive = urge + goal situation
Physiological needs

- Thirst
- Hunger
- Rest
- Warmth
- Libido
- ...

→ Survival as emergent property
Social needs

- Affiliation (Attention from others, external legitimacy)
- Internal legitimacy
- Nurturing (caring for others)
- Affection
- Dominance
Cognitive needs

• Competence:
  – Skill acquisition (epistemic competence)
  – Coping/control ability (general competence)
  – Effect generation

• Uncertainty reduction:
  – Exploration

• Aesthetics:
  – Stimulus oriented
  – Structure oriented (abstract aesthetics)
Needs and urges

Physiological Needs
- Sustenance
- Pain avoidance
- Rest
- Libido
- (...and many more)

Social Needs
- Affiliation
- Nurturing
- Affection
- Legitimacy
- Dominance

Cognitive Needs
- Competence
- Exploration
- Aesthetics

Target Value
Current Value
Need

Urge Strength
Pleasure
Displeasure
Urgency

12/2/15 FutureAI
Motivational Learning

• association by learning:

\[ \Delta s_i \rightarrow \text{urge indicator} \]

\[ s_i \rightarrow \text{demand} \]

\[ \Delta s_i \rightarrow \text{change indicator} \]

\[ V^+ \rightarrow A^+ \]

\[ V^- \rightarrow A^- \]

\[ w_1 \]

\[ w_2 \]

\[ \text{valence} \rightarrow \text{associator} \]

\[ \text{goal situation} \]

\[ \text{aversive situation} \]
Motivational Learning

- retrogradient reinforcement

Protocol Chain

- Demand
- Urge Indicator
- Change Indicator
- Valence
- Associator

\[ \Delta s_i \]

\[ w_1 \]

Goal
Motivational Learning

Motivator:

situations leading up to goal = plan

\[ w_1 \]

urge

\[ S_i \]

autonomous regulation

\[ \Delta S_i \]

valence

\[ V^+ \]

associator

\[ A^+ \]

goal
Intention:

Motivational Learning

$S_i$ 

urge 

autonomous regulation 

goal 1 

goal 2 

goal 3 

... 

goal $n$
Motivational learning
Motive selection

Need becomes active

No autonomous regulation possible:
Trigger Urge Signal

Try to satisfy urge opportunistically

No opportunistic satisfaction possible:
Urge Strength – Suppression > Strength of Leading Motive:
Try to recall strategy to satisfy urge

motive strength = \frac{\text{expected reward} \times \text{urgency} \times \text{competence}}{\text{cost of strategy}}

If no strategy is found:
Construct a plan to satisfy urge

If no plan is found:
Increase need for exploration

Turn strongest motive into leading motive (intention)
Need parameters

• Strength: relative importance
• Decay: rate of replenishment
• Gain: effect of satisfaction
• Loss: effect of frustration

• different configuration of need parameters = different personality traits
Modulation in PSI/MicroPsi
Primary modulators

- Arousal: unspecific sympathicus syndrome
- Valence: situation evaluation (good/bad)
- Aggression: fight or flight
Compare: Affective dimensions (Wundt 1910)
Attentional modulators

- Resolution Level: width of focus
- Supression: depth of focus; motive stability
- Securing Rate: rate of checking the environment

Diagram:
- Detailed Cognition/Perception
- Narrow Focus
- Internal
- Securing Rate
- External
- Wide Focus
- Fast Cognition/Perception
Modulator dynamics

- Resolution Level
- Arousal
- Supression
- Securing Rate
- Aggression
- Urgency: All Needs
- Urgency: Leading Motive
- Exploration Need (Uncertainty)
- Obstacle Prevents Satisfaction
- Urge Strength: All Needs
- Urge Strength: Leading Motive
- General Competence
- Competence for Current Task
Modulator parameters

- Baseline
- Range
- Volatility
- Duration

- Different modulator parameter configurations = different temperaments
Emotions as directed affect + Modulation

Examples:

**Fear:** anticipation of aversive events ($\rightarrow$ neg. valence) + arousal

**Anxiety:** uncertainty ($\rightarrow$ neg. valence) + low competence + arousal, high securing behavior (frequent background checks)
**Emotions as directed affect + Modulation**

Examples:

**Anger:** Perceived obstacle (usually agent) manifestly prevented reaching of an active, motivationally relevant goal (→ neg. valence), sanctioning behavior tendency (→ goal relevance is re-directed to sanctioning of obstacle), arousal, low resolution level, high action readiness, high selection threshold

**Sadness:** Manifest prevention from all conceived ways of reaching active, relevant goal, without relevant obstacle (→ neg. valence), support-seeking behavior (by increased demand for affiliation), low arousal, inhibition of active goal → decreased action readiness
Emotions as directed affect + Modulation

Examples:

• **Pride**: high competence (→ low securing rate), high internal legitimacy, likely coincidence with high external legitimacy

• **Joy**: high arousal + high perceived reward signal from satisfying a demand

• **Bliss**: low arousal + high perceived reward signal from satisfying a demand (since physiological demands often involve high arousal, mostly related to cognitive demands, such as aesthetics)
Individual Variations by Parameterizing

Possible grounding of personality properties (FFM):

- **Openness**: appreciation of art and new ideas, curiosity
- **Conscientiousness**: rulefollowing vs. chaotic
- **Extraversion**: tendency to seek stimulation by environment and others
- **Agreeableness**: tendency for cooperativeness and compassion
- **Neuroticism**: emotional stability, effect of failure to self-confidence
Needs and Big Five

- Affiliation
- Nurturing
- Affection
- Legitimacy
- Dominance
- Competence
- Exploration
- Aesthetics
- Openness
- Conscientiousness
- Extraversion
- Agreeableness
- Neuroticism
Player personality types

Motivation and personality

- Personality properties can be modeled as motivational variability

![Motivation and personality diagram](image)
Needs and player types (with S. Tekovsky)
Motivation in MicroPsi

- All actions are directed on goals or avoidance of aversive goals
- All goals are established through learning how to satisfy needs
- All decisions are based on strengths of urges and chances to satisfy corresponding needs
- Personality differences are the result of parametrization of the motivational system
Emotion in MicroPsi

- Affective states are configurations of cognition, by modulators
- Primary modulators: arousal, valence, aggression
- Attentional modulators: focus, securing rate, resolution level
- High-level emotions are determined by an affective state that is directed on motivational content