Introduction to the MicroPsi Architecture

Joscha Bach
joscha@mit.edu
Overview

In this introduction:

- Requirements for AGI architectures
- Representation and symbol grounding
- Configurable cognition: understanding emotion
- Motivation and autonomy
MicroPsi Architecture

• Origins: Psi theory (Dörner et al. 1999, 2002)

Dietrich Dörner
Universität Bamberg
MicroPsi Architecture

• Origins: Psi theory (Dörner et al. 1999, 2002)
  - Framework for simulating cognitive agents
  - Explorative agents in virtual worlds
  - Evolutionary simulations
  - Learning and classification
  - Robot control
• MicroPsi 2 (2012–current):
  - new simulation framework
  - adaptation for knowledge representation
Basic premises

- Are minds characterized by a single organizing principle, or by a complex set of very particular constraints?

→ Human minds are solution to a very specific class of control problems, with generally applicable intelligence as a by-product.
Cognitive Artificial Intelligence

Methods should focus on components and performances necessary for intelligence:

• **Whole, testable architectures**

• **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**
Components for Cognitive AI

- Universal mental representations
  (compositional + distributed $\rightarrow$ neurosymbolic)
Components for Cognitive AI

• (Semi-) General problem solving: Operations over these representations

(neural learning, categorization, planning, reflection, consolidation, ...)

![Diagram of cognitive operations](image)
Components for Cognitive AI

- Perceptual grounding
Components for Cognitive AI

- Perceptual grounding and action

Diagram showing the integration of components for cognitive AI, including perception and action.
Components for Cognitive AI

- Perceptual grounding and action
Components for Cognitive AI

- Model of current situation, and protocol of past situations

![Diagram](image-url)

- Perception
- Protocol memory
- Situation model
- Action

- $\text{Op}_1$, $\text{Op}_2$, $\text{Op}_3$, $\text{Op}_4$, $\ldots$, $\text{Op}_n$
Components for Cognitive AI

- Model of self

![Diagram showing components of self-model](image)
Components for Cognitive AI

- Abstractions of objects, episodes and types
Components for Cognitive AI

• Anticipation of future developments

![Diagram showing components of cognitive AI]

- Self model
- Plans
- Perception
- Action

Components for Cognitve AI

1. Op₁
2. Op₂
3. Op₃
4. Op₄
... n

Declarative memory
Procedural memory
Frame
Expect.
Components for Cognitive AI

- Action selection and executive control

![Diagram showing the components of cognitive AI]

- Longterm memory
- World model
- Mental stage

- Perception
- Action
Components for Cognitive AI

- Action selection and executive control

```
Longterm memory     World model     Mental stage
```

```
Perception ---> Action selection ---> Action
```

Cognitive Integration  MicroPsi
Components for Cognitive AI

- Universal motivation: autonomous identification of goals

![Diagram of Cognitive AI components]

- Perception
- Action selection
- Action
- Longterm memory
- World model
- Mental stage
- Motivational system: Motive Selection and Decision making
Components for Cognitive AI

- Emotional modulation and affect
Components for Cognitive AI

- Whole, testable architectures

Principles of Synthetic Intelligence
(Bach 2003, 2009)
“Classical Cognitive Architectures” tend to focus on cognition as an isolated problem solving capability.
MicroPsi Architecture—simplified
MicroPsi Architecture—simplified

MicroPSI Agent

Meta-Management

Behaviour Script Space / Execution Space

Internal Behaviours

Motivation
Execution

Long Term Memory

Short Term Memory / Local Perceptual Space

Memory Maintenance

Sensors/Modulators

Motivation / Emotion

Cognitive Functions

Body Parameters

Perception

Action

External Behaviors
Cognitive urges:
- Competency
- Uncertainty reduction

“physiological” urges:
- Food, intactness

Motive selection mechanism
MicroPsi Principles: Emotion

[Emotional] modulators

Modulation of planning and action control

Modulation and urge parameters define basal emotional states
**MicroPsi Principles: Perception**

- **Model of environment**
- **Primitive sensor data**
MicroPsi Principles: Perception

Interpretation by hypotheses *(Hypercept)*

Model of environment

Primitive sensor data
**MicroPsi Principles: Action**

![Diagram of MicroPsi Principles: Action]

- **Execution**
- **Goals and current plans**
- **External Behaviors**
- **Internal Behaviors**
- **Motivation**
- **Execution**
- **Short Term Memory/Local Perceptual Space**
- **Long Term Memory (LTM)**
- **Meta-Management**
- **Body Parameters**
- **Perception**
- **Action**

---

**MicroPsi Agent**

---

Cognitive Integration

MicroPsi

28
Agent Functionality

- Episodic learning
- Goal-directed behavior, motivational system
- Emotional modulation
- Hypothesis based perception
- Simple planning
- Execution of hierarchical plans
Implementation: MicroPsi (Bach 03, 05, 04, 06)
Implementation: MicroPsi (Bach 03, 05, 04, 06)
Implementation: MicroPsi (Bach 03, 05, 04, 06)
Implementation: MicroPsi (Bach 03, 05, 04, 06)
Implementation: MicroPsi (Bach 03, 04, 05, 06)

Robot control

Low-level perception

Multi-agent interaction

Control and simulation
Different simulation environments
Implementation: MicroPsi 2 (Bach, Welland 12)
MicroPsi 2

- New software basis (more lightweight and modular)
- Minecraft as (optional) simulation platform (implementation: Kemper 2014)
- Currently: design of social games for testing of motivation model
Representation in Psi: Neurons and Quads
"Cortex" and Activators
Spreading Activation in Quads
Spreading Activation in Quads
Spreading Activation in Quads
Spreading Activation in Quads

"Cortex"

Activator por  Activator ret  Activator sub  Activator sur
Spreading Activation in Quads
Spreading Activation in Quads

"Cortex"

Activator
por
Activator
ret
Activator
sub
Activator
sur
Spreading Activation in Quads
Spreading Activation in Quads
Spreading Activation in Quads
Spreading Activation in Quads
Spreading Activation in Quads
From Quads to Node Nets

→ “Quad”
Node Nets: Concept Nodes

Basic Building Block: **Concept Node**

- **Concept**
  - id
  - gen
  - por
  - ret
  - sub
  - sur
  - cat
  - exp
  - sym
  - ref

- **w, certainty, (x,y,z,t)**

- General activation
- Is-followed-by/causes
- Follows-after/is-caused-by
- Has-part/has-attribute
- Is-part-of/attributes
- Class
- Member
- Symbol
- Refers-to
Node Nets: Link Types

- **SUB:** has-part/has-attribute
- **SUR:** part-of
- **POR:** leads-to/causes
- **RET:** comes-from
- **SYM:**
- **EXP:**
- **CAT:** is-a
- **REF:** refers-to
Node Nets: Concept Nodes

Basic Building Block: Concept Node

Gates:
- gate function
- gate parameters

Slots

Node activation function; state
Building agents with node nets

• Graphical representation + python shell
• Node types:
  – individual NN elements (sigmoidal neurons etc.)
  – concept nodes
  – script nodes (state machines for distributed execution)
  – sensors and actuators
  – node spaces
  – control nodes
  – native modules
  – new: multi-state nodes
Request Confirmation Networks

[Diagram of Request Confirmation Networks]

Cognitive Integration

MicroPsi

56
Using hierarchies for perception

**HyPercept** (Hypothesis based Perception, Dörner 1999)

- Features from environment trigger sensors
- Sensors activate object hypotheses (activation spreads upwards)
- Object hypotheses are successively tested (activation spreads downwards)
- Sensors confirm or invalidate hypotheses
- Finally:
  - If only one hypothesis left: successful recognition
  - If multiple hypotheses left: further exploration
  - If no hypothesis left: new object
Hierarchical Representation
Protocol Memory

diagram showing the flow of information from episodic schema, behavior program, to situation and action, with pathways to other behavior programs and urge indicator.
Why is emotion so interesting?

• Obviously: applications
• *Phenomenal aspect* of emotion – *feeling*:
  Emotion may be a critical juncture when it comes to understanding the mind

→ What is emotion?
Emotion

• Having an emotion is different from behaving as if having an emotion

• What is it like to have an emotion?

• Can emotion only be simulated, or can an artificial system be in an emotional state?

→ is having an emotion a way or an aspect of information processing?
Modeling emotion

Most models of emotion account for description:
- What emotional states can a system have?
- Which events (in the environment) trigger emotions?
- How can emotional states be expressed?
- How can emotional states be recognized?

(Plutchik 80; Ortony, Clore & Collins 88; Scherer 93; Gratch & Marsella 04)

vs. “How can a system have an emotion?” (functional)
Conceptual analysis: Ortony, Clore, Collins 1988:

Valenced reaction to

- consequences of events
  - pleased, displeased etc.
  - focusing on consequences for self
    - prospect relevant
      - joy
    - prospect irrelevant
      - distress
  - focusing on consequences for other
    - desirable for other
      - happy for...
    - undesirable for other
      - resentment
gloating
pity

- actions of agents
  - approving, disapproving etc.
  - focusing on self (agent)
    - pride
    - shame
    - reproach
  - focusing on other agent
    - admiration
    - hatred

- aspects of objects
  - liking, disliking, etc.

confirmed disconfirmed

- satisfaction
  - relief
  - gratification
  - gratitude
  - prospect-based
  - well-being/attribution compounds

- prospect-based compounds

- well-being/attribution compounds
  - gratification
  - gratitude
  - remorse
  - anger
Affective dimensions in the PSI theory (Dörner 1999)

- **Valence** (appetence/aversion)
- **Arousal** (unspecific sympathicus syndrome)
- **Selection threshold** (motive dominance)
Compare: Affective dimensions (Wundt 1910)
Affective dimensions in the PSI theory

- **Valence** (appetence/aversion)
- **Selection threshold** (motive dominance)
- **Arousal** (unspecific sympathicus syndrome)
- **Resolution level** (focus)
- **Goal directedness**
- **Securing rate**
The Psi theory about emotion

• Affect is seen as a configuration of a cognitive system

• Modulators of cognition:
  – arousal, selection threshold, securing threshold, resolution level
  – estimate of competence and certainty
  – pleasure/distress signals → valence

• Affective state is emergent property of modulation

• Directed affects (higher-level emotions) emerge by association of demand with appetive or aversive objects/situations
Purpose of emotional modulation

- Control width, depth and bias of operations on mental representations of the agent
  → modify perception, memory, planning and action selection
- Reduce complexity of cognitive processes
Motivational System

- All goals attempt to satisfy a (hard-wired) demand → flexible goals, but (evolutionary) suitable behavior
Motivational System

- Drives correspond to set of demands of the agent

<table>
<thead>
<tr>
<th>Food</th>
<th>Water</th>
<th>Integrity</th>
<th>Affiliation</th>
<th>Internal Legitimacy</th>
<th>General and Specific Competence</th>
<th>Uncertainty Reduction</th>
<th>Evolutionary and Abstract Aesthetics</th>
</tr>
</thead>
</table>

- Physiological
- Social
- Cognitive
Physiological Drives

• if autonomous regulation of body processes fails
  → actively manage physiology (seek food, water, healing, shelter, rest, warmth, ...)
  → escape perilous situations
  → implicitly seek physical survival
Social Drives

• **Affiliation**: structure social interaction beyond rational utility

• increased by ‘legitimacy signals’, decreased by ‘anti legitimacy signals’ (and adaptively over time); allows for non-material reward and punishment

• *external* legitimacy: group acceptance

• *internal* legitimacy: “honor”, conformance to internalized social norms
Cognitive Drives

- Competence
  - epistemic (problem specific)
  - general (ability to satisfy demands)
  - effect oriented

- Uncertainty reduction
  - novelty seeking

- Aesthetics
  - evolutionary preferences (stimulus oriented)
  - abstract (representation oriented)
**Motivational System**

All possible goals correspond to (at least one) demand

<table>
<thead>
<tr>
<th>Food</th>
<th>Water</th>
<th>Integrity</th>
<th>Affiliation</th>
<th>Internal Legitimacy</th>
<th>General and Specific Competence</th>
<th>Uncertainty Reduction</th>
<th>Evolutionary and Abstract Aesthetics</th>
</tr>
</thead>
</table>

Physiological | Social | Cognitive
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
- Urge Strength
- Pleasure
- Displeasure
- Urgency

Cognitive Integration MicroPsi
Motivational System

- drive = demand + urge indicator
Motivational Learning

• motive = urge + goal situation
Motivational Learning

• motive = urge + goal situation

Water

target

current level

urge indicator

$s_{\text{water}}$

goal situation

aversive situation
Motivational Learning

- association by learning:

\[
\begin{align*}
S_i & \quad \text{urge indicator} \\
\Delta S_i & \quad \text{change indicator} \\
V^+ & \rightarrow A^+ \\
V^- & \rightarrow A^- \\
\text{valence} & \quad \text{associator} \\
goal \text{ situation} & \\
\text{aversive situation}
\end{align*}
\]
Motivational Learning

- retrogradient reinforcement

![Diagram showing a protocol chain leading to a goal with nodes for urge indicator, demand, change indicator, valence, and associator.]
Motivational Learning

Motivator:

situations leading up to goal = plan

\[ \Delta s_i \]

\[ V^+ \]

\[ A^+ \]

Valence

Associator

autonomous regulation

urge

\[ w_1 \]
**Motivational Learning**

Intention:

- Goal 1
- Goal 2
- Goal 3
- ... (omitted)
- Goal n

Diagram:

- Autonomous regulation
- Urge
- $S_i$
Motivational Learning

- association by learning:

\[ S_i \quad \Delta S_i \quad V^+ \quad A^+ \quad V^- \quad A^- \quad w_1 \quad w_2 \]

- urge indicator
- demand
- change indicator
- goal situation
- aversive situation
- valence
- associator

Cognitive Integration  MicroPsi
**Motive Selection**

A motive is raised to an intention based on:

- relative strength of urge: 
  \[ \max(s_1, s_2, s_3, ..., s_n) \]
- opportunity
- expected success probability (heuristics) \( \rightarrow \) value/risk
- selection threshold (adaptive; urgency)

![Diagram](image-url)
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

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

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

If no plan is found:
Increase need for exploration

Turn strongest motive into *leading motive* (intention)
Motivation and Modulator Dynamics

- Urgency:
  - Leading Motive
  - All Motives

- Importance:
  - Leading Motive
  - All Motives

- Resolution Level
- Arousal
- Goal Directedness
- Selection Threshold
- Securing Behavior
- Experienced Uncertainty
- Competence:
  - Specific to current task
  - General
Emotions as directed affect + Modulation

Examples:

**Fear:** anticipation of aversive events (→ neg. valence) + arousal

**Anxiety:** uncertainty (→ 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 readyness, 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 readyness
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)
Example: computational creativity

- Creativity involves exploration of space of possible solutions to an incrementally defined problem.

Diagram:

1. Identify (partial) constraints
2. Generate candidates
3. Evaluate against constraints
4. Repair/adapt
5. Elaborate
Example: computational creativity

• Creativity as a learning problem:
  – optimize problem representation
  – optimize constraint identification
  – optimize solution traversal

• Source of reward signal:
  – abstract aesthetics (better representations)
  – uncertainty reduction (problem space exploration)
  – competence (specific problem solving skills)

• “Eureka” moment: large “positive” delta during constraint match → reward signal
Example: Mapping to FFM (Big Five)

Demand dynamics:

- Food
- Water
- Integrity
- Affiliation
- Internal Legitimacy
- Gen./Epis. Competence
- Uncertainty Reduction
- Aesthetics

Physiological
Social
Cognitive
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
Example: Mapping to FFM (Big Five)

Demand dynamics:

Physiological    Social    Cognitive
Example: Mapping to FFM (Big Five)

Demand dynamics:
Example: Mapping to FFM (Big Five)

Valence: Pleasure/Pain signals
Example: Mapping to FFM (Big Five)

**Neuroticism**: stronger experience of negative emotions, lower emotional stability

(strong negative reward/stronger loss of competence, certainty; stronger decay -> more frequent need for replenishment, possibly lower gain), proneness to anxiety due to loss of certainty
**Example: Mapping to FFM (Big Five)**

**Extraversion**: surgency, activity in social relations, expressivity
(strong gain for affiliation and competence, high decay of affiliation)
**Example: Mapping to FFM (Big Five)**

**Openness**: desire for novelty, intellectual independence, non-conservatism, appreciation for art and new ideas

(Strong gain for uncertainty reduction, high epistemic competence for uncertainty reduction, probably strong rewards for all demands)
Example: Mapping to FFM (Big Five)

**Agreeableness:**

(strong positive and negative reward for affiliation, lower gains for competence)
Example: Mapping to FFM (Big Five)

**Conscientiousness**, Rigidity:
(high loss in competence, high selection threshold, possibly low effect of demands on arousal and resolution level)
Thank you!

- Project home: micropsi.com
- Questions, ideas: joscha@mit.edu