

# A Framework for Self-Awareness in Artificial Subjects

Axel Jantsch<sup>1</sup>   Kalle Tammemäe<sup>2</sup>

<sup>1</sup>Vienna University of Technology  
Vienna, Austria

<sup>2</sup>Tallinn University of Technology  
Tallinn, Estonia

International Conference on Hardware-Software Codesign  
and System Synthesis, 2014

## Sometimes Machines Behave Silly



## Sometimes Machines Behave Silly



# Sometimes Machines Behave Silly



# Sometimes Machines Behave Silly



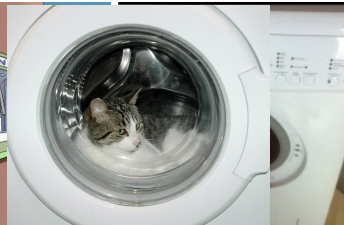
# Sometimes Machines Behave Silly



# Sometimes Machines Behave Silly



# Sometimes Machines Behave Silly





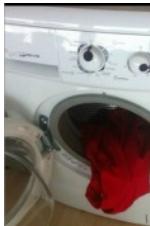
# Sometimes Machines Behave Silly



# Sometimes Machines Behave Silly



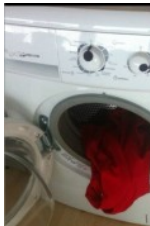
# Sometimes Machines Behave Silly



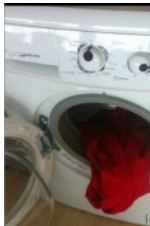
# Sometimes Machines Behave Silly



# Sometimes Machines Behave Silly



# Sometimes Machines Behave Silly



## The Benefits of Awareness

- ▶ Better functionality in different contexts
- ▶ Context depending performance
- ▶ Appropriate reaction in presence of faults



## The Benefits of Awareness

- ▶ Better functionality in different contexts
- ▶ Context depending performance
- ▶ Appropriate reaction in presence of faults



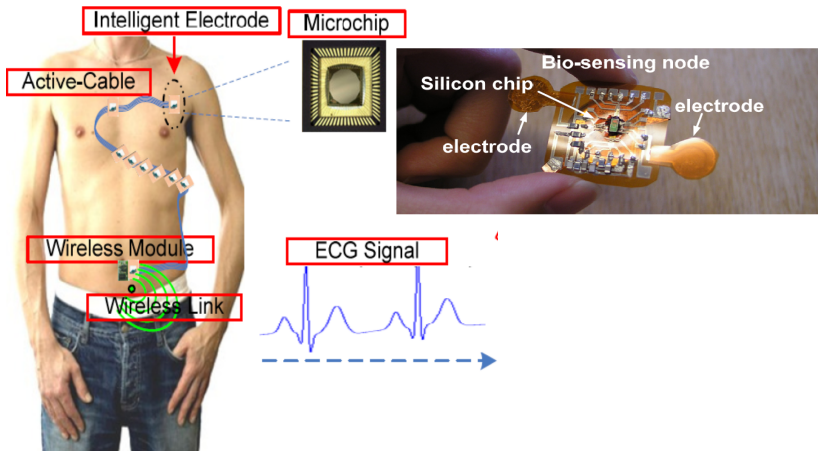


## The Benefits of Awareness

- ▶ Better functionality in different contexts
- ▶ Context depending performance
- ▶ Appropriate reaction in presence of faults



# Self-Awareness for Resource Constrained, Insect-like Gadgets



## Properties of Awareness

- ▶ Not all information is necessary
- ▶ More information does not imply more awareness

- ▶ Raw data is interpreted/abstracted
- ▶ Data interpretation is “meaningful”
- ▶ The drawn conclusions are “robust”
- ▶ The reaction is appropriate

## Properties of Awareness

- ▶ Not all information is necessary
- ▶ More information does not imply more awareness

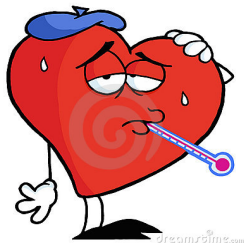
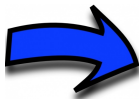
- ▶ Raw data is interpreted/abstracted
- ▶ Data interpretation is “meaningful”
- ▶ The drawn conclusions are “robust”
- ▶ The reaction is appropriate

## BioPatch: Temperature Sensor

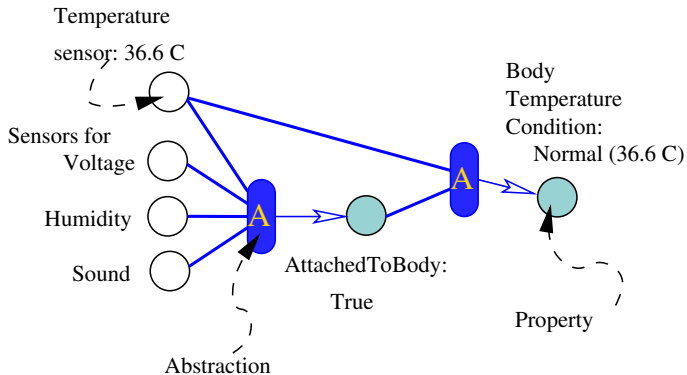
measured temperature	{	$< 20$	→ person is dead
		in $[20, 32]$	→ alive, life threatening
		in $[32, 36]$	→ worrying, not life threatening
		in $[36, 37]$	→ normal
		in $[37, 37.5]$	→ elevated, not worrying
		in $[37.5, 39.5]$	→ fever
		in $[39.5, 43]$	→ high fever, life threatening
		$> 43$	→ person is dead

## Abstractions and Models

Abstraction: Mapping of Measurements  $\Rightarrow$  Properties

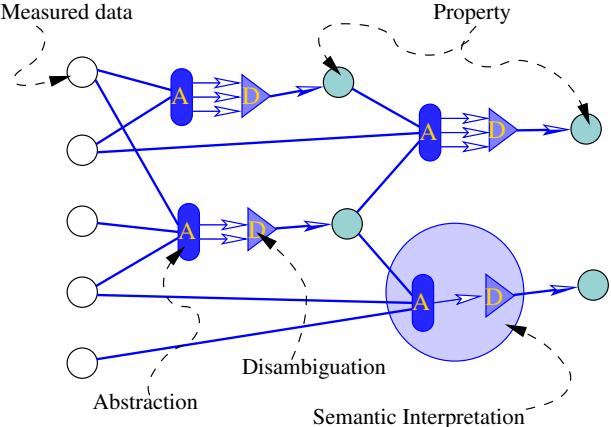


## Abstractions and Models



# Disambiguation

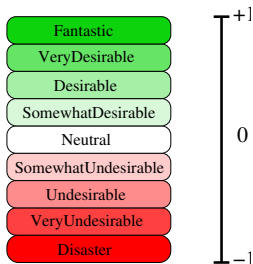
## Selection among several interpretations





## Desirability Scale

A value range that captures the desirability of something



**Semantic Attribution** maps the values of a property to a point in the desirability scale.

## History

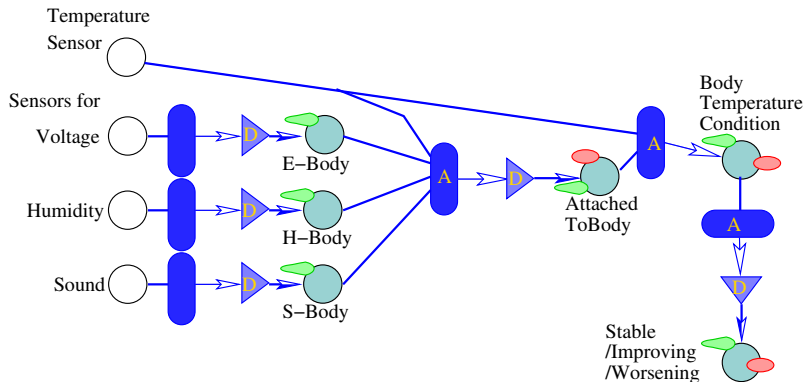
**History of a Property** The evolution of the values of a property.

**Abstracted History** The history stores abstracted values.

**Attributed History** The history is annotated with attributions.

**Fading History** If the property values are more abstracted the longer ago they have occurred.

## Sensors and properties of the BioPatch



## Expectations

### Expectation on Environment

- ▶ all implicit and explicit assumptions about the environment;
- ▶ a value range for each of the monitored properties.

### Expectation on Subject

- ▶ all implicit and explicit assumptions about the subject;
- ▶ a value range for each of its monitored properties.

## Goals

**Sub-Goal** A sub-goal of the subject is a desired value range of a property of the subject or its environment.

**Goal** A goal consists of one or several sub-goals.

**Purpose** The purpose of a subject is to achieve all its defined goals.

## Inspection and Simulation

**Self Inspection Engine** is a mapping from a set of properties onto a desirability scale;

**Model Transformation** Given a model and a set of actions, a transformation applies actions and derives the new values for all properties.

**Simulation** Given a model and a set of potential actions, a simulation is a sequence of transformations applied onto the model resulting in a new, updated model.

## Awareness of a Property

- ▶ The subject makes observations and derives the property by means of a meaningful semantic interpretation (*Meaning Condition*).
- ▶ The semantic interpretation is robust (*Robustness Condition*).
- ▶ There is a semantic attribution which is meaningful (*Attribution Condition*).
- ▶ The subject reacts appropriately to its perception of the property (*Appropriateness Condition*).
- ▶ A history of the evolution of the property over time is maintained (*History Condition*).

## Awareness of a Subject

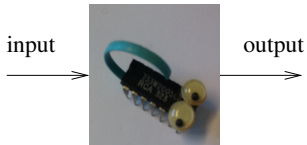
- ▶ The subject can assess how well it meets all its goals (*Goal Condition*).
  
- ▶ The subject can assess how well the goals are achieved over time and when its performance is improving or deteriorating (*Goal History Condition*).



## Levels of Awareness

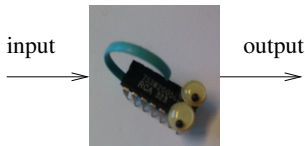
## Levels of Awareness

**Level 0 - Functional:** Behaviour is an immediate function of input.

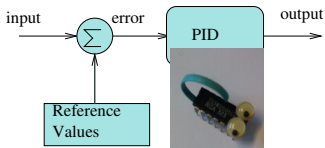


## Levels of Awareness

**Level 0 - Functional:** Behaviour is an immediate function of input.

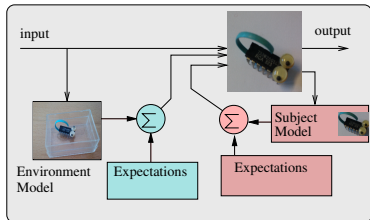


**Level 1 - Adaptive:** Output is an adaptive reaction to the input and a reference value (PID controller).



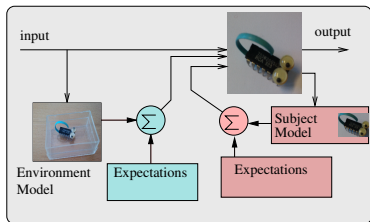
## Levels of Awareness

**Level 2 - Self-aware:** System represents some of its own properties and its environment as an abstraction. The models are related to desirable reference points.

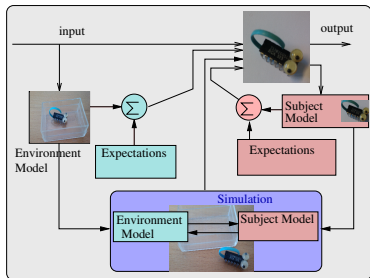


## Levels of Awareness

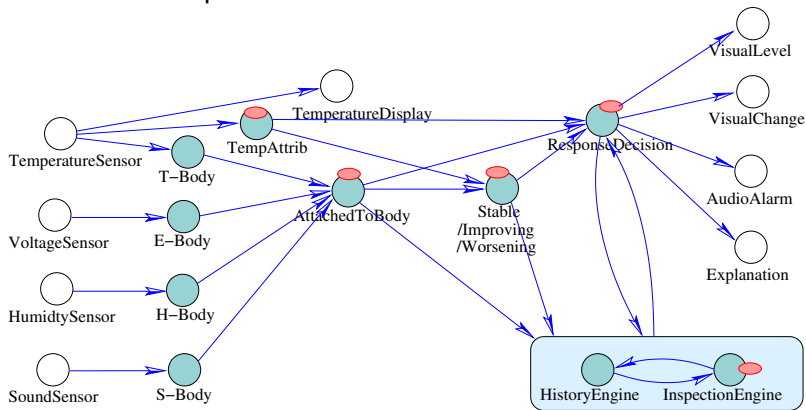
**Level 2 - Self-aware:** System represents some of its own properties and its environment as an abstraction. The models are related to desirable reference points.



**Level 3 - Predictive:** System can simulate the effect of future input and of its own actions on the Self-Rep and the environment.



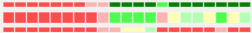




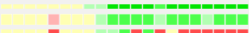
# BioPatch Example



# BioPatch Example

BioPatch

Scenario DownUp2

Temperature:	38.9	
Level:		
Temp Change:	Decreasing	
Attached:	Most likely at Body (0.8)	
Audio Alarm:	No Alarm	
Explanation:		Normal
History Attribution:		
Quit		

## Summary

- ▶ Awareness and self-awareness are useful properties
  - ▶ Context dependent functionality
  - ▶ Context dependent performance
  - ▶ Appropriate behavior in all situations
- ▶ Necessary features:
  - ▶ Data abstraction
  - ▶ Disambiguation
  - ▶ Desirability mapping
  - ▶ History maintenance
  - ▶ Expectations and goals
  - ▶ Self-inspection
  - ▶ Prediction and simulation

### Challenges:

- ▶ Application specific selection and tuning of features
- ▶ Online learning and adaptation
- ▶ Efficient implementation



## Summary

- ▶ Awareness and self-awareness are useful properties
  - ▶ Context dependent functionality
  - ▶ Context dependent performance
  - ▶ Appropriate behavior in all situations
- ▶ Necessary features:
  - ▶ Data abstraction
  - ▶ Disambiguation
  - ▶ Desirability mapping
  - ▶ History maintenance
  - ▶ Expectations and goals
  - ▶ Self-inspection
  - ▶ Prediction and simulation

### Challenges:

- ▶ Application specific selection and tuning of features
- ▶ Online learning and adaptation
- ▶ Efficient implementation