A Framework of Awareness for Artificial Subjects

Axel Jantsch
Vienna University of Technology, Austria
axel.jantsch@tuwien.ac.at

Kalle Tammemäe
Tallinn University of Technology, Estonia
kalle.tammemae@ttu.ee

1. INTRODUCTION

A small battery driven bio-patch, attached to the human body and monitoring various vital signals such as temperature, humidity, heart activity, muscle and brain activity, is an example of a highly resource constrained system, that has the demanding task to assess correctly the state of the monitored subject (healthy, normal, weak, ill, improving, worsening, etc.), and its own capabilities (attached to subject, working sensors, sufficient energy supply, etc.). These systems and many other systems would benefit from a sense of itself and its environment to improve robustness and sensibility of its behavior. Although we can get inspiration from fields like neuroscience, robotics, AI, and control theory, the tight resource and energy constraints imply that we have to understand accurately what technique leads to a particular feature of awareness, how it contributes to improved behavior, and how it can be implemented cost-efficiently in hardware or software. We review the concepts of *environment-* and *self-models, semantic interpretation, semantic attribution, history, goals and expectations, prediction,* and *self-inspection,* how they contribute to awareness and self-awareness, and how they contribute to improved robustness and sensibility of behavior.

Researchers have for some time realized that a sense of "awareness" of many embedded systems' own situation is a facilitator for robust and dependable behaviour even under radical environmental changes and drastically diminished capabilities. This insight has recently led to a proliferation of work on self-awareness and other system properties such as self-organization, self-configuration, self-optimization, self-protection, self-healing, etc., which are sometimes subsumed under the term "self-*".

Self-awareness is often treated as a means to, or as a special feature of adaptive or autonomic computing. A more conceptual approach to awareness by J-S. Preden is rooted in awareness theories for humans and the situation awareness techniques for human-machine interfaces, that were developed to present humans with context sensitive information [9]. Corbato and coworkers have developed the classic concept of control into self-aware and conscious control [3], by using a model to represent the controller itself, and a meta-controller that can modify the mode and parameters of the plant controller based on the controller model, its goals and performance.

For instance, L. Guang describes an agent-based system for self-aware embedded application [7], and in a recent book [2] various approaches to self-monitoring, self-configuration, self-optimization and self-adaption are described. Self-awareness is often treated as a means to, or as a special feature of adaptive or autonomic computing.

However, if we compare devices described or conceived in these papers with naturally found awareness, we realize quickly that there are profound differences. Even though we associate awareness or even consciousness with many animals such as apes, monkeys, dogs or cats, hardly anybody would attribute the same quality of awareness to the power manager proposed by L. Guang [7] or to the Heartbeats performance monitor described by M. D. Santambrogio et al. [10].

A number of theories of self-awareness, consciousness and attention in human brains have been developed. Baars describes the *Global Workspace Model* (GWM) [1] where only exactly one of the many parallel, subconscious processing modules can get access at any given time and thus controlling the activation in large parts of the brain. Dehaene and coworkers have developed GWM further and refined it into a *Global Neural Workspace* (GNW) hypothesis [6]. In computer simulations of neural networks many of the phenomena could be confirmed, that are also observed in experiments with humans [5]. Francis C. Crick, and Christof Koch [4] hypothesis integrated conscious role is played by sheet-like structure called claustrum, located deep in inner surface of neocortex. Jeff Hawkins postulates a basic learning algorithm uniformly active throughout the neocortex that specializes different regions based on the different input signals they receive [8]. Hawkins' *Hierarchical Temporal Memory* (HTM) has recently gained wide attention from many research groups that have diligently studied and implemented HTM in several variations and with various purposes.

So, assuming that awareness has potential advantages for many, simple or complex Cyber-Physical Systems, and given several hypothesis of the operation and emergence of awareness in humans and animals, we ask the following question: How can we realize and implement awareness efficiently so that it becomes feasible and useful in resource constraint systems such as small an inexpensive bio-sensors or artificial, flying insects?

ESWEEK'14 October 12-17 2014 New Delhi, India

http://dx.doi.org/10.1145/2656075.2661644.

2. DEFINITIONS

An aware CPS, which we call *Subject*, can be aware of itself and it can be aware of its Environment, or both. Obviously, it can only partially monitor itself and its environment, and usually only a tiny fraction of the potentially useful data is collected. However, complete knowledge of the environment is not necessary and the subject can still be aware of certain aspects. Hence, awareness is not a linear function of information volume. Even the collection of huge amounts of data does not make the Subject aware of its environment. On the other hand, a small amount of data can be used to generate awareness if it is *abstracted* and *interpreted* sensibly.

We are inclined to call the Subject aware of a certain aspect of the Environment, if three conditions are met:

(i) The data interpretation is meaningful;
(ii) The drawn conclusions are robust; and
(iii) The reaction of the device is appropriate.

A few terms and concepts constitute a framework of awareness of artificial devices, that allows for an accurate distinction between levels of awareness and different capabilities related to awareness.

Abstraction Given a set of measurement data $\mathcal{M}$, a set of properties $\mathcal{P}$, and a property $P \in \mathcal{P}$, an abstraction defines the mapping of measured data and properties to values of the P.

Disambiguation If an abstraction of a set of measurements and properties leads to more than one interpretation, disambiguation selects one of them and assesses the certainty of the interpretation.

Semantic Interpretation Abstraction and Disambiguation together is called a Semantic Interpretation of a set of measurement data $\mathcal{M}$ and properties $\mathcal{P}$ leading to a new property P that has to be meaningful in a given context.

Desirability Scale A value range that captures the desirability of something from "very undesirable" to "very desirable" is called a Desirability Scale DS.

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

History of a Property The evolution of the values of a property is called its history. A plain history is factual, i.e. all the measured or abstracted values are stored. Alternatively, in case of fading history, more distant values are averaged, thus keeping the amount of required memory reasonable.

Goal A goal consists of one or several sub-goals. We denote the set of goals that refer to subject properties as $\mathcal{G}_S$, and those goals that deal with environment properties as $\mathcal{G}_E$.

Purpose The purpose of a subject is to achieve all its defined goals to an as high degree as possible.

Expectation on Environment The expectations of a subject in its environment $\mathbf{E}_E$ consists of all implicit and explicit assumptions about the environment necessary for the subject to operate properly.

Expectation on Subject The expectations of the subjects in itself $\mathbf{E}_S$ consists of all implicit and explicit assumptions about the subject necessary for it to operate properly.

Inspection Engine The inspection engine is a mapping from a set of properties $\mathcal{P}$ onto a desirability scale DS, i.e. it inspects all the properties and assesses them by deriving a value on the desirability scale.

3. AWARENESS AND SELF-AWARENESS

We are now in a position to formulate seven conditions for awareness.

Awareness of a Property Given a property P, we distinguish the following condition for being aware of property P:

(C.1) The subject makes physical measurements or observations that are used to derive the values of P by means of a meaningful semantic interpretation (*Meaning Condition*).
(C.2) The semantic interpretation is robust (*Robustness Condition*).
(C.3) There is a semantic attribution which is meaningful (*Attribution Condition*).
(C.4) The subject's reaction to its perception of P is appropriate (*Appropriateness Condition*).
(C.5) A history of the evolution of the property over time is maintained, in particular of the increasing or decreasing deviations over time (*History Condition*).

Awareness of a Subject For a subject to be aware of itself, it must relate to its goals and understand how well it meets them.

(C.6) The subject can assess how well it meets all its goals, thus it has an understanding which goals should be achieved and to which extent they are achieved (*Goal Condition*).
(C.7) The subject can assess how well the goals are achieved over time and when its performance is improving or deteriorating (*Goal History Condition.*

Depending on how fully aware a subject is of itself and its environment, we distinguish between five levels of awareness.

Awareness Level 0 A functional subject instinctively reacts to a given input always in the same manner; its output is a mathematical function of its input. If it fulfills the conditions (C.1)-(C.4) we call it aware at level 0.

Awareness Level 1 An adaptive subject tries to minimize the difference between input values and corresponding reference values by using a proportional-integral-derivative (PID) controller or a similar algorithm. If it meets conditions (C.1)-(C.4) it is aware at level 1.

Awareness Level 2 A self-aware subject

1. is aware of at least one subject property and one environment property according to conditions (C.1)-(C.4) and condition (C.6),
2. it contains an inspection engine that periodically derives one integrated attribution of the subject as a whole, and
3. it computes its actions based on
   (a) the monitored and attributed properties $\mathcal{P}_S$ of the subject and of the environment $\mathcal{P}_E$;
   (b) the attributed expectations $\mathbf{E}_S$ on the subject and on the environment $\mathbf{E}_E$;
   (c) the set of goals $\mathcal{G}_S$ and $\mathcal{G}_E$

Awareness Level 3 A history sensitive self-aware subject fulfills all requirements of level 2 and, in addition, fulfills the history conditions (C.5) and (C.7)

Awareness Level 4 A predictive subject is a history sensitive self-aware subject of level 3 and, in addition, its decision making process involves a simulation engine, that can simulate the effects of actions on the environment and on the subject, thereby predicting future states and behaviors of both the subject and its environment. In case of contradiction between predicted and measured state (anomalous situation) system starts to seek for the best match according to current situation (shifting focus to alternative set of goals). The simulation engine selects the simulation scenarios and then decides the actions to be taken based on predictions of the simulations.

Awareness Level 5 In addition to self-awareness, *group awareness* means that the subject distinguishes between itself, the environment and the *peer group*. The latter is treated differently by associating it with peer group specific expectations and goals.

The framework sketched above is based on a set of concepts that (a) seem to be practically useful for robust, reliable and autonomous behaviour, and (b) are inspired by the phenomena of awareness and consciousness in animals and humans. Although there is broad and strong evidence for the utility of these concepts, future work has to demonstrate how to efficiently implement them for scalability under tight resource constraints.

4. REFERENCES

[1] B. Baars. *A Cognitive Theory of Consciousness.* Cambridge University Press, 1989.

[2] P. Cong-Vinh, editor. *Autonomic Networking-on-Chip: Bio-Inspired Specification, Development, and Verification.* CRC Press, 12 2011.

[3] C. H. Corbato. *Model-based Self-awareness Patterns for Autonomy.* PhD thesis, Universidad Polytécnica de Madrid, Madrid, Spain, 2013.

[4] F. C. Crick and C. Koch. What is the function of the claustrum? *"Philosophical Transactions of the Royal Society"*, pages 1271–1279, 2005.

[5] S. Dehaene and J.-P. Changaeux. Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2):200–227, 4 2011.

[6] S. Dehaene, S. Kerszberg, and J.-P. Changaeux. A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Acadademy of Science of the USA*, 95(24):14529–14534., 11 1998.

[7] L. Guang. *Hierarchical Agent-based Adaptation for Self-Aware Embedded Computing Systems.* PhD thesis, University of Turku, Turku, Finland, 2012.

[8] J. Hawkins and S. Blakeslee. *On Intelligence.* St. Martin's Griffin, New York, 2004.

[9] J.-S. Preden. *Enhancing Situation-Awareness, Cognition and Reasoning of Ad-Hoc Network Agents.* PhD thesis, Tallinn University of Technology, 6 2012. Thesis on informatics and system engineering C56.

[10] M. Santambrogio, H. Hoffmann, J. Eastep, and A. Agarwal. Enabling technologies for self-aware adaptive systems. In *NASA/ESA Conference on Adaptive Hardware and Systems (AHS)*, pages 149–156, 2010.