Awareness in Autonomic Systems

General Properties

Short-/Long-term Impact

Open Issues
Outline

• General properties
  – Self-awareness
  – Perception
  – Collectivity
  – Internal model

• Short-/long-term impact
  – Safety
  – Sustenability
  – Ethical and philosophical

• Open issues
Some related general properties

DIFFERENT LEVELS OF AWARENESS
Neisser's levels of self-awareness

1. **Ecological self**
   (Awareness of internal or external stimuli).

2. **Interpersonal self**
   (Awareness of interactions with others).

3. **Extended self**
   (Awareness of time: past and/or future).

4. **Private self**
   (Awareness of one's own thoughts, feelings, intentions).

5. **Conceptual self**
   (Awareness of one's own self-awareness, possession of an abstract model of oneself).

See Neisser (1997).

The conceptual self has the capacity for “meta-self-awareness”, being aware that one is self-aware.

Emergence of self-awareness

- In collective systems, the entire system can appear self-aware,
- Though constituent parts may exhibit less self-awareness themselves.
- Self-information is distributed about the system, and not present at any single point.
- See Mitchell (2005).

Computational framework

We would like to take some of these ideas, and translate and apply them to the design of computing systems.

Why?

• Provide a common understanding and language for self-aware computing.
• Relate computing concepts to psychological basis – draw inspiration from natural systems.
• Enable the principled engineering of self-aware systems by identifying common features and how to build them.
Levels of Computational Self-awareness

- Ecological self → Stimulus awareness
- Interpersonal self → Interaction awareness
- Extended self → Time awareness
- Private self → Goal awareness
- Conceptual self → Meta-self-awareness
Multi-level self-aware systems

“Self” is a concept, not a box.
Emergent self-awareness: implications for system design

• Systems can exhibit behaviour which appears globally self-aware,
• No single component is required to possess system-wide self-knowledge.
• Need not require that a self-aware system possesses a global controller!
• Sufficient for components just to have local knowledge, of relevant parts.
Computational self-awareness

• To be self-aware, a system should:
  ▪ Possess knowledge of its internal state (*private self-awareness*),
  ▪ Possess knowledge about its environment (*public self-awareness*).

• Optionally, it might also:
  ▪ Possess knowledge of its interactions with others and the wider system (*interaction awareness*),
  ▪ Possess knowledge of time, e.g. past and likely future experiences (*time awareness*),
  ▪ Possess knowledge of its goals e.g. objectives, preferences, constraints (*goal awareness*),
  ▪ Select what is and is not relevant knowledge (*meta-self-awareness*).
Computational self-awareness capabilities

• Where systems differ in terms of their self-awareness, is in what knowledge is available and collected, and how it is represented.

• Key questions of a system:
  ▪ Which level(s) of self-awareness are present?
  ▪ How are its self-awareness capabilities implemented?
Some related general properties

PERCEPTION
Awareness requires perception

• Perception is extracting the relevant information from the environment and from itself in order to be able to act appropriately.

• Perception is a difficult task as beings are surrounded by a lot of information and data.

• Perceiving the relevant information depends on the context and the purpose of a task.

• Thus there is a subtle interplay between awareness and perception.
Awareness requires perception (2)

- Perception is a complicated process that requires appropriate sensing mechanisms.
- Perception can require forms of memory, knowledge and learning.
- Thus, perception can involve complicated forms of cognition.
- Awareness and perception allow producing appropriate attention.
Awareness requires perception (3)

• Appropriate attention depends on what you are.
• Each type of intelligent machine and each individual machines can require different appropriate attention.
• Appropriate attention is complicated because it cannot be simply directly programmed
  it has to emerge from complex interactions between the individuals, their environment, the context, the tasks, their current states, their history, etc.
Some related general properties

COLLECTIVITY/SWARM/DISTRIBUTEDNESS
“Natural” Complex Systems

All agent types: molecules, cells, animals, humans & tech

- The brain
- Organisms
- Ant trails
- Termite mounds
- Animal flocks
- Internet, Web
- Markets, economy
- Social networks
- Cities, population
- Molecules
- Cells
- Animals
- Humans & tech

All agent types: molecules, cells, animals, humans & tech

Awareness

Self-Awareness in Autonomic Systems

Slides Factory
“Natural” Complex Systems

Natural and human-caused categories of complex systems

... yet, even human-caused systems are “natural” in the sense of their unplanned, spontaneous emergence.
“Natural” Complex Systems

Architectured natural complex systems (without architects)

- biology strikingly demonstrates the possibility of combining pure self-organization and elaborate architecture

- the brain
- organisms
- ant trails
- termite mounds
- living cell
- physical patterns
- Internet, Web
- markets, economy
- social networks
- animal flocks
- cities, populations

Awareness

SlidesFactory
“Natural” Complex Systems

Emergence on multiple levels of self-organization

complex systems:

a) a large number of elementary agents interacting locally

b) simple individual behaviors creating a complex emergent collective behavior

c) decentralized dynamics: no master blueprint or grand architect
From genotype to phenotype, via development
From cells to pattern formation, via reaction-diffusion
From social insects to swarm intelligence, via stigmergy
From birds to collective motion, via flocking
From neurons to brain, via neural development.
Common Properties of Complex Systems

• **Emergence**
  – the system has properties that the elements do not have
  – these properties cannot be easily inferred or deduced
  – different properties can emerge from the same elements

• **Self-organization**
  – “order” of the system increases without external intervention
  – originates purely from interactions among the agents (possibly via environment)

• **Positive feedback, circularity**
  – creation of structure by amplification of fluctuations
    ▪ ex: the media talk about what is currently talked about in the media

• **Decentralization**
  – the “invisible hand”: order without a leader
    ▪ **distribution**: each agent carry a small piece of the global information
    ▪ **ignorance**: agents don’t have explicit group-level knowledge/goals
    ▪ **parallelism**: agents act simultaneously
Spontaneous Self-Organization of Human-Made Systems

• Burst to large scale: *de facto* complexification of ICT systems
  – ineluctable breakup into, and **proliferation** of, modules/components

→ trying to keep the lid on complexity won’t work in these systems:
  - cannot place every part anymore
  - cannot foresee every event anymore
  - cannot control every process anymore ... but do we still *want* to?
Spontaneous Self-Organization of Human Organizations

• Burst to large scale: *de facto* complexification of organizations, via techno-social networks
  – ubiquitous ICT capabilities connect people and infrastructure in unprecedented ways
  – giving rise to complex techno-social systems composed of a multitude of human users and computing devices
  – explosion in size and complexity in all domains of society:
    ▪ healthcare  ▪ energy & environment
    ▪ education  ▪ defense & security
    ▪ business  ▪ finance
  – large-scale systems have grown and reached unanticipated levels of complexity, beyond their components’ architects

→ impossible to assign every single participant a predetermined role
The Need for Computational Models

- **ABM** meets **MAS**: two (slightly) different perspectives

**CS science:** understand “natural” CS → Agent-Based Modeling (ABM)

**CS engineering:** design a new generation of “artificial” CS → Multi-Agent Systems (MAS)

-but again, don’t take this distinction too seriously! they overlap a lot
Regaining Control of Self-Organization

- ... by exporting models of natural CS to ICT: “(bio-)inspired” engineering

**CS Science:** observing and understanding "natural", spontaneous emergence (including human-caused)

<table>
<thead>
<tr>
<th>ex: neurons &amp; brain</th>
<th>ex: ant colonies / bird flocks</th>
<th>ex: genes &amp; evolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>biological neural models</td>
<td>trails, swarms / collective motion</td>
<td>laws of genetics</td>
</tr>
<tr>
<td>binary neuron, linear synapse</td>
<td>move, deposit, follow “pheromone” / separation, alignment, cohesion (“boids”)</td>
<td>genetic program, binary code, mutation</td>
</tr>
<tr>
<td>artificial neural networks (ANNs) applied to machine learning &amp; classification</td>
<td>ant colony optimization (ACO) graph theoretic &amp; networking problems / particle swarm optimization (PSO) “flying over” solutions in high-D spaces</td>
<td>genetic algorithms (GAs), &amp; evolutionary computation for search &amp; optimization</td>
</tr>
</tbody>
</table>

**CS (ICT) Engineering:** creating and programming a new, artificial self-organization / emergence
Some related general properties

INTERNAL MODEL
Internal Models

• A characteristic of all (?) self-aware systems is that they have internal models

• What is an internal model?
  – It is a mechanism for representing both the system itself and its current environment
    – example: a robot with a simulation of itself and its currently perceived environment, inside itself
  – The mechanism might be centralized (as in the example above), distributed, or emergent
Internal Models

• Why do self-aware systems need internal models?
  – Because the self-aware system can run the internal model and therefore test what-if hypotheses*
    • what if I carry out action x..?  
    • of several possible next actions, which should I choose?
  – Because an internal model (of itself) provides the self in self-aware

*Reference: Dennett’s model of ‘generate and test’
Examples

• Examples of conventional internal models, i.e.
  – Analytical or computational models of *plant* in classical control systems
  – Adaptive connectionist models such as online learning Artificial Neural Networks (ANNs) within control systems
  – GOFAI symbolic representation systems

• Note that internal models are *not* a new idea
Examples 1

- A robot using self-simulation to plan a safe route with incomplete knowledge

Examples 2

• A robot with an internal model that can learn how to control itself

Examples 3

• ECCE-Robot
  – A robot with a complex body uses an internal model as a ‘functional imagination’

Examples 4

- A distributed system in which each robot has an internal model of itself and the whole system
  - Robot controllers and the internal simulator are co-evolved

A Generic Architecture

• The major building blocks and their connections:

- Sense data
- Control System
- Actuator demands
- Internal Model

The IM moderates action-selection in the controller

The IM is initialized to match the current real situation

The loop of generate and test evaluates the consequences of each possible next action
Short-/Long-term impact

SAFETY
The safety problem

• For any engineered system to be trusted, it must be safe
  – We already have many examples of complex engineered systems that are trusted; passenger airliners, for instance
  – These systems are trusted because they are designed, built, verified and operated to *very stringent* design and safety standards
  – The same will need to apply to autonomous systems
The safety problem (2)

• The problem of safe autonomous systems in *unstructured* or *unpredictable* environments, i.e.
  – *robots* designed to share human workspaces and physically interact with humans *must* be safe,
  – yet guaranteeing safe behaviour is extremely difficult because the robot’s human-centred working environment is, *by definition*, unpredictable
  – it becomes even more difficult if the robot is also capable of *learning* or *adaptation*
Safety

• No system can have pre-determined responses to every eventuality in unpredictable environments
  • example: robots that have to interact with humans
    – therefore no system that works in unpredictable environments can be guaranteed to be safe
  – Self-awareness could provide a powerful solution to this fundamental problem
Safety

• How can a self-aware system be safer (than a system without self-awareness)?
  – Because a self-aware system with an internal model of itself and its environment could*
    1. *Represent* the currently perceived (unforeseen) situation in its internal model
    2. *Run* each possible next action in its internal model (in a sense *imagine* each course of action)
    3. *Evaluate* the *safety* of each action
    4. *Choose* the *safest* of those actions, and then actually carry out that action

*a major engineering challenge is to build a system that can do this quickly
Short-/Long-term impact

SUSTAINABILITY
Sustainable Futures

- Make Critical infrastructure more adaptive
  - Royal Commission on Environmental Pollution
  - Tragedy of the Commons not inevitable
- Take into account
  - Social arrangements of citizens
  - Attributes of the infrastructure with which the interact
  - Context of institutions
Sustainable Futures (2)

• Adaptive Institutions
  – Individuals, ICT-enabled devices and institutions are deeply entangled
  – ICT devices can be equipped with social awareness and can participate in the collective endeavour
  – Out of the entanglement new structures can emerge
  – People retain the power to self-organise these structures

• Computational Sustainability
  – There is a reason why Elinor Ostrom won the Nobel Prize for Economic Science – empowering individuals with collective awareness
Short-/Long-term impact

PHILOSOPHICAL
Philosophy

- The conception and implementation of self-aware systems might have philosophical implications
  - If self-aware systems are, in some way, models of living systems then could we gain insights into self-awareness in living systems by testing such models?
  - Is self-awareness the first step toward long-term goals of artificial theory-of-mind, and machine consciousness?
  - Could we gain *ontological* insights by asking questions such as, at what point does a self-aware system make the transition to a self-determining autonomous agent, i.e. ‘being’
Could a robot be ethical?

• An ethical robot would require:
  – The ability to *predict the consequences* of its own actions (or inaction)
  – A set of *ethical rules* against which to test each possible action/consequence, so it can choose the most ethical action
  – New legal status..?
Using internal models

• *Internal models* might provide a level of *functional self-awareness*
  – sufficient to allow robots to ask *what-if* questions about both the consequences of its next possible actions
  – the same internal modelling architecture could conceivably embody *both safety and* ethical rules
Open Issues

QUESTIONS & CHALLENGES
Research questions and challenges

• Dilemma of wishing to make our designed artefacts autonomous but not too much (safety).
• To have a metrics to measure properties related to awareness, autonomy.
• We do not know how to engineer self-organization and emergence.
• We do not know how to cope with autonomy and variability. Dilemma of system stability and reliability incorporating randomness and variability.
• How to design and implement self-aware systems?
• What kind of tools and methodology can we use here?
• Is it ethical to build self-aware systems?
• Can we build autonomic self-aware systems that behave in an ethical way? Related: legally correct behaviour, behaviour compliant with some set of rules and regulations.
• What makes known natural systems self-aware?
• Describing the scope of the future behaviour of a self-aware system.
• Predicting the behaviour of autonomic systems and their interactions with the environment.
Research questions and challenges

• How to ensure safety and security of autonomic self-aware systems? How to differentiate malicious from benign behaviour?
• What does the system theory of autonomic self-aware systems look like?
• How to build an autonomic self-aware system that would last 100 years?
• To what extent can Big Data be treated as an autonomic self-aware system?
• Can you separate an autonomic self-aware system from its environment?
• In what sense is human and machine self-awareness different? What implications do these differences have on developing them?
• How can we draw inspiration from human self-awareness for designing machine self-awareness?
• How to do the second order design needed in autonomic self-aware systems?
• Will autonomic self-aware systems develop their own medical science?
• Goal: build an autonomic self-aware energy production system.
• Goal: build a smart city / computer network / communication network.
Acknowledgment

The slides in this presentation were produced with contributions from:

Peter Lewis
Rene Doursat
Jose Halloy
Alan Winfield