A Survey of Self-Awareness and Its Application in Computing Systems

Peter R. Lewis*, Arjun Chandra†, Shaun Parsons*, Edward Robinson*, Kyrre Glette†,
Rami Bahsoon*, Jim Torresen† and Xin Yao*

* School of Computer Science
University of Birmingham, UK
{p.r.lewis—ug39sxp—e.r.robinson—r.bahsoon—x.yao}@cs.bham.ac.uk
† Department of Informatics
University of Oslo, Norway
{chandra—kyrrehg—jimtoer}@ifi.uio.no

Abstract—**Novel computing systems are increasingly being composed of large numbers of heterogeneous components, each with potentially different goals or local perspectives, and connected in networks which change over time. Management of such systems quickly becomes infeasible for humans. As such, future computing systems should be able to achieve advanced levels of autonomous behaviour. In this context, the system's ability to be self-aware and be able to self-express becomes important. This paper surveys definitions and current understanding of self-awareness and self-expression in biology and cognitive science. Subsequently, previous efforts to apply these concepts to computing systems are described. This has enabled the development of novel working definitions for self-awareness and self-expression within the context of computing systems.**

Keywords-**self-awareness; self-expression; autonomous systems.**

I. INTRODUCTION

As the development of computing systems continues, they increasingly comprise large numbers of heterogeneous components, each with potentially different goals or local perspectives, and connected in networks with dynamic topologies. Management of such systems quickly becomes infeasible for humans. As such, future computing systems, from robots to personal music devices to web services, should be able to achieve advanced levels of autonomous behaviour, in order to manage and adapt themselves. Nevertheless, users still expect high performance, reliability, security and other qualities.

In order for a system to effectively manage itself and adapt to changing circumstances, its ability to be self-aware becomes important. Self-awareness is concerned with the availability, collection and representation of knowledge about a system, by that system. The presence of such knowledge permits reasoning and intelligent decision making, which can support effective, autonomous adaptive behaviour.

In Section II, this paper surveys current definitions and understanding of self-awareness in biology and cognitive science, which provides a basis for the consideration of how the concept of self-awareness might translate to computing. Previous efforts to apply these concepts to computing

systems are described in Section III, which enables the development of a novel working definition for self-awareness within the context of computing systems; this is described in Section IV.

While self-awareness is concerned with a system's knowledge about itself, we have found it useful to separate this self-awareness from the process of determining a system's resulting actions. This enables us to explicitly consider these two aspects separately; the latter we term *self-expression*. Section V presents this idea and a similar working definition for computing systems that are self-expressive. Section VI concludes the paper.

In realising self-awareness and self-expression in computing systems, contributions from many disciplines will be required; amongst them are psychology, philosophy, economics, complexity science, artificial and computational intelligence and electronic and software engineering. This paper presents steps towards establishing a methodology for engineering systems that are both self-aware and self-expressive.

II. SELF-AWARENESS: INSPIRATION FROM BIOLOGY AND COGNITIVE SCIENCE

The study of self-awareness emerged as a field within psychology in the 1960's, despite being first discussed in much earlier literature [1], [2]. Distinct from, but building upon consciousness, Morin defines self-awareness as *"the capacity to become the object of one's own attention"* [3], and a self-aware organism as one that *"becomes aware that it is awake and actually experiencing specific mental events"* [3]. As a prerequisite for this, the organism must have the capability to monitor or observe itself.

A. Public and Private Self-Awareness

Duval and Wicklund have written extensively on self-awareness [4], defining two classes of self-awareness: *subjective* and *objective*. Objective self-awareness is described as being *"focused exclusively upon the self and consequently the individual attends to his conscious state, his personal history, his body, or any other personal aspects of himself"*

[4]. Subjective self-awareness by contrast is described as *"a state of consciousness in which attention is focused on events external to the individuals consciousness, personal history, or body"* [4].

Similarly, Goukens et al. [5] describe a distinction between *public* and *private* self-awareness in humans. As with objective self-awareness, private self-awareness is concerned with internal knowledge known only to the individual. Externally unobservable, an individual's knowledge might include, for example, being hungry or having a headache. More complex private knowledge might include the individual's values, opinions or goals. Goukens et al.'s public self-awareness, as with subjective self-awareness, is more concerned with how the individual may be perceived externally. This might include how the individual appears to others, its social relationships or the observation of its effect on the environment.

In summary, public self-awareness is concerned with knowledge about the environment in which the individual finds itself, including knowledge concerning consequences of the individual's own actions. This requires the ability to observe the environment or others, and is a reflective process. Private self-awareness is by contrast concerned with knowledge solely concerning the individual. This knowledge is typically not available to others, unless the individual communicates it. An organism which is fully self-aware, in both the public and private senses, must therefore have the capability to possess knowledge of the external environment and its internal state.

B. Levels of Self-Awareness

In addition to the public and private classes of self-awareness, Neisser [6] developed a model describing five levels of self-awareness:

1) **Ecological self**
   The ecological self is the most minimal form of self-awareness. It provides only for basic stimulus-response interaction, as the organism has a basic awareness of stimuli. The ecological self can be thought of as the minimum requirement for the organism to not be unconscious.
2) **Interpersonal self**
   The interpersonal self enables the organism to possess a simple awareness of its external interactions, permitting limited adaptation to others in the performance of tasks.
3) **Extended self**
   The extended self extends the interpersonal self to permit reflection of interactions over time. The organism is aware of the existence of past and future interactions.
4) **Private self**
   The private self includes that the individual can process more advanced information concerning itself,

such as thoughts, feelings and intentions.
5) **Conceptual self**
   The conceptual self (or self-concept) is the most advanced form of self-awareness, representing that the organism is capable of constructing and reasoning about an abstract symbolic representation of itself.

This final, most advanced level of self-awareness also permits what is sometimes termed **meta-self-awareness** [7]. This is an awareness on the part of the organism that it is itself self-aware. Meta-self-awareness permits complex reasoning and analysis of both public and private self-awareness processes. Examples of meta-self-awareness might include that the organism is aware that it is angry about something, or that it has been feeling emotional recently.

C. The Emergent Appearance of Self-Awareness

So far, we have just considered self-awareness in the context of a single organism. However, biological systems often consist of collectives of organisms or cells, which engage socially, perhaps in symbiotic or parasitic relationships. Considering the brain, Mitchell [8] notes that it appears from an external perspective to be self-aware. This awareness, she describes as:

> *"Information about the global state of the system, which feeds back to adaptively control the actions of the system's low-level components. This **information about the global state is distributed** and statistical in nature, and thus is difficult for observers to tease out. However, the system's components are able, collectively, to use this information in such a way that **the entire system appears** to have a coherent and useful sense of its own state."* [8]

The emphasis here has been added, to highlight that a system which behaves in a self-aware manner is not necessarily required to possess a single component which has access to system global knowledge. Indeed, in many cases, e.g., ant colonies, immune systems and humans themselves, the entire system *appears* self-aware, despite the knowledge available at constituent parts being only local. The appearance of self-awareness is an emergent effect.

This is a key observation which can contribute to the design of self-aware systems: one need not require that such a system possesses a global omniscient controller. Indeed, many natural systems appear to have been favoured by evolution which do not have such a central point of control, and rely upon relevant knowledge being available at required locations within the system. It is highly likely that this can improve the robustness and adaptability of such systems; these are desirable properties for natural and artificial systems alike.

III. PREVIOUS WORK IN SELF-AWARE COMPUTING

Various research initiatives have used the term *self-awareness* explicitly as a property of their computing machinery within computer science and engineering. From these initiatives a number of clusters stand out as significant efforts to incorporate self-awareness within computing systems. These clusters will be briefly described in the remainder of this section.

A. Meta-Cognition

The higher levels of self-awareness, such as meta-self-awareness introduced in section II-B, can also be viewed as meta-cognition, defined [9] as knowing about knowing. Cox argues that being aware of the self is not merely possessing information, but being able to use that information in order to generate goals, which may lead to the information being modified. Importantly, Cox [10] also suggests that meta-cognition is similar to the algorithm selection problem, wherein the task is to choose the most efficient algorithm from a set of possibilities. This has much in common with the conceptual self discussed in section II-B.

Integration of intelligence into systems such that they are self-aware in the meta-cognitive form has been on DARPA's research agenda for some time [11]. Architectural issues in building such integrated systems which then exhibit self-awareness have also been considered and discussed during a DARPA workshop [12].

B. Engineering Self-Aware Systems

While meta-cognition or meta-self-awareness are concerned with higher reasoning abilities in AI, at a more fundamental level, efforts exist to engineer systems which explicitly consider knowledge about themselves. Agarwal et al. [13], [14] put forward a case for a paradigm shift in system design practice. The idea here is to move from a procedural design methodology wherein the behaviour of the computing system is pre-programmed or considered beforehand (i.e., at design time), towards a self-aware system where this is not required and the system adapts to its context at run-time. One aim is to avoid or reduce the need to consider the availability of resources and various other constraints beforehand, instead intelligently trading-off available resources for performance at run-time.

A recent example [15] of the application of such a design approach is the use of intelligent agents to control the behaviour of distributed smart cameras, which are given the goal of tracking objects that move through their field of view. Here, agents use a utility function to represent their progress towards the tracking goal. In attempting to maximise their utility, they make intelligent decisions at run-time concerning how to exchange tracking responsibilities between themselves and with whom to communicate. The emergent result is an efficient outcome in the trade-off between tracking performance and communication overhead.

Unlike previous approaches to this problem, the cameras did not require any *a priori* knowledge of their environment or the camera neighbourhood structure.

Importantly, for a system to be self-aware it is not required to be highly complex; indeed the scalability of the concept means that self-awareness has also been considered in much simpler systems. An example of this is so called *cognitive radio devices* [16], which monitor and control their own capabilities and also communicate with other radio devices to monitor theirs. This enables them to improve the efficiency of communication by negotiating changes in parameter settings [17].

Agarwal et al. [13] argue that five design properties should be considered when engineering self-aware systems. Namely, that they are:

- introspective, i.e. they can observe and optimise their own behaviour,
- adaptive, i.e. they can adapt to changing needs of applications running on them,
- self-healing, i.e. they can take corrective action if faults appear whilst monitoring resources,
- goal oriented, i.e. they attempt to meet user application goals, and
- approximate, i.e. they can automatically choose the level of precision needed for a task to be accomplished.

C. Self-Awareness in Pervasive Computing

This work is primarily concerned with systems that are mobile and hence their context changes. As such, they need to monitor their own state and the external environment they are in, in order to adapt to changes in the environment in a context specific way. Often monitoring and adaptation are studied in the context of human-computer interaction, since the interest is on how such systems self-adapt in order to be useful to humans in different situations (e.g., "going for a run"). A recent survey [18] covers issues and challenges involved in assimilating sensor data from a myriad of sources in order for pervasive computing systems to identify situations which human users may be in. Ye et al. show a shift in techniques over time from *logic based* ones towards those that are *learning based*, as the sensor data has become more complex, erroneous and uncertain, with sensors becoming ever more pervasive. The learning of mappings between sensor data and the situation, given current model building techniques, poses challenges such as the lack of training data, which can lead to low performing models. This has been tackled by considering unsupervised learning [19], [20] and web mining [21], allowing for extracting common sense knowledge. Another line of research within pervasive computing concerns constructing simulation models of contexts, for context-aware applications to be tested in [22].

D. Systems with Emergent Self-Awareness

Self-awareness research is not limited to an entity or system in itself being able to monitor and reason about itself, but also describes emergent phenomena [8]. In natural systems like ant colonies and the immune system, the awareness of the global state is distributed across the elementary units that make up the system (e.g. ants and their trails) and is statistical in nature. This helps the system stay robust at the global level in the face of disturbances. In essence, the system as a whole is aware enough of itself to understand when the globally stable state gets disturbed, and engages the elementary units to collect information locally, which builds up in a statistical fashion, helping the elementary units use this statistical information to get the system back into the globally stable state. It has been proposed [8] that such systems can provide guidelines for designing artificial intelligence systems with decentralised architectures, for example robotic swarms, which exhibit apparent self-awareness.

One example of where such a system has been developed is within the SWARM-BOTS project [23]. One of the objectives of the SWARM-BOTS project was the design and implementation of a novel mobile robot, called an *s-bot*. While s-bots' individual capabilities within an environment are physically limited (much like individual insects in the natural world), through local communication they are able to *self-assemble* [24] into larger structures, known as *swarm-bots*, which are capable of achieving goals not reachable by individual s-bots. Examples of such goals might include navigation over challenging terrain or the transportation of large objects; in all cases, these tasks cannot be solved without the coordinated movements of individual *s-bots*.

Other research challenges within the area of self-assembly, or *structural self-organisation*, include better understanding how system structures, rather than individual behaviours, can be adapted over time with respect to the system's distributed sense of *self-concept* [25].

E. Formal Approaches to Self-Awareness

More formal approaches to achieving self-awareness and self-expression in complex systems are also being carried out. For example, the ASCENS project seeks to simplify the building of such systems by applying formal methods from software engineering. Vassev and Hinchey [26] discuss in some detail the current state of the art in knowledge representation within these types of systems, and lay out both formal deterministic and probabilistic approaches for representing knowledge. They argue that this will facilitate better self-awareness through easier analysis of the states and goals of local parts of the system. The project's focus is currently on swarms (called *ensembles*) of autonomous robots (called *service components*), but is relevant to any application discussed in this paper.

IV. WORKING DEFINITION FOR SELF-AWARE COMPUTING SYSTEMS

As we have seen, the term self-awareness is used in a variety of different ways within computer science and engineering literature. The general concept of self-aware computing covers but is not limited to all of these cases. Indeed, given the disparate and occasionally overly rigid use of the term self-awareness in the literature, we have found it useful to develop a more general working definition of a self-aware computing system, drawing on the biological inspiration discussed in Section II.

This definition is based on the idea of a conceptual component called a *self-aware node*. A node in this context need not physically exist as a hardware or software component of a computing system, but provides a conceptualisation of locality within a global system, particularly in relation to what is considered *self* in the context of self-awareness. This distributed nature of conceptual components is particularly relevant to the idea of distributed self-awareness, as expounded by Mitchell [8]. The definition is as follows.

To be self-aware a node must:
- *Possess information about its internal state (private self-awareness).*
- *Possess sufficient knowledge of its environment to determine how it is perceived by other parts of the system (public self-awareness).*

Optionally, it might also:
- *Possess knowledge of its role or importance within the wider system.*
- *Possess knowledge about the likely effect of potential future actions / decisions.*
- *Possess historical knowledge.*
- *Select what is relevant knowledge and what is not.*

Given the description of self-awareness discussed in Section II, if a node possesses only public self-awareness then it would only be able to access knowledge of other nodes within the system or the environment the node is operating within. Conversely, a node which possesses only private self-awareness would have no knowledge of other nodes or the operating environment, but would instead have knowledge about itself: perhaps its state, current behaviour or history.

Possession of both public and private self-awareness allow these two sources of knowledge to be combined to provide a meaningful context for adaptation and behavioural decisions. This knowledge will, for example, be able to support both simple reactive behaviour as well as complex learning, prediction and action selection tasks.

V. SELF-EXPRESSION

As we have seen, self-awareness is concerned with knowledge about a system by that system, which may be centrally held or distributed in nature. However, in designing

autonomous self-aware systems, we have found it useful to explicitly and separately consider the process of determining a system's actions as a result of this knowledge. This process we call self-expression.

A. Self-Expression in Psychology

The term self-expression is less frequently and less ambiguously defined in the literature, when compared with self-awareness. Dictionary definitions offer a rather intuitive understanding of self-expression: *"the expression of ones feelings, thoughts, or ideas..."* according to the Oxford English Dictionary[1], while Merriam-Webster[2] offers *"the assertion of ones individual traits"*. The latter has been used in social psychology literature [27]. Similarly, the psychologists Chen et al. [28] define a self-expressive individual as one which behaves *"in line with their states and traits."* This is consistent with the dictionary definitions.

Furthermore, Chen et al. [28] claim that the relative power of a person affects the amount of self-expression which the person engages in. Individuals in positions of social power are found to express their thoughts, feelings and attitudes more than those with relatively less power. This observation from human societies may provide a useful mechanism to be examined in computational scenarios.

B. Self-Expressive Computing

Despite the lack of previous treatment of the idea of self-expression in the computing and engineering literature, we have developed a working definition similarly to our definition for self-expressive computing. Recalling from the definition above, a self-expressive individual's behaviour is determined by its feelings, thoughts and ideas. Clearly, a computing node does not have feelings, thoughts or ideas, as such, but it does have knowledge about itself, such as its state, context, goals, values, objectives and constraints, which could be considered analogous. For self-expressive computing capabilities, Section II also discussed that the relative power of a person in a social situation has an effect on the magnitude of their self-expression (i.e. the degree to which they assert their behaviour). Translating this idea to the computing domain, this could be interpreted in such a way that the magnitude of the behaviour expressed by a given node is proportional to a notion of authority in the network. However, we can widen the idea beyond a simple scale of authority, to consider that nodes' behaviour may be determined by their role within the wider system. This provides further motivation for a node to be aware of such a role and to build this into its knowledge of its context.

From the definitions discussed above we can define two properties that define a self-expressive computing node.

- *A node exhibits self-expression if it is able to assert its behaviour upon either itself or other nodes.*

[1] *Oxford dictionaries online:* http://oxforddictionaries.com
[2] *Merriam-Webster:* http://merriam-webster.com

- *This behaviour is based upon the node's state, context, goals, values, objectives and constraints.*

An example of the benefit of considering the self-awareness and the self-expressive properties of a system separately can be provided in the context of the distributed smart camera system [15] described in section III. Here, individual cameras within a decentralised network are self-aware, in that they collect and process information about their state and context, such as the currently visible objects given their position, the confidence associated with their tracking them and knowledge of their relationship to neighbouring cameras in the network. However they are also self-expressive; they make decisions about which objects to track and how to bid to neighbouring cameras for tracking rights. Specifically, their communication strategy determines how to balance the trade-off between communication and performance by selectively targeting cameras based on historical knowledge. Thus, the self-awareness informs the self-expression of the camera. Clearly, the processes associated with knowledge and those with actions must both be attended to in optimising the cameras' design.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented the concept of self-awareness as understood in biology and cognitive science. Subsequently, we have discussed previous and ongoing efforts to incorporate this concept into computing systems, identifying several clusters of research effort. Based on the biological literature, we have presented a novel working definition of self-awareness in a computing system, which provides a general concept and permits scalability from simple to highly complex systems, as well as direct or emergent self-awareness properties.

In considering a system's actions as a result of its self-awareness, we presented the separate but related concept of self-expression: the process of determining a system's actions as a result of its knowledge of itself. This permits the distinction of a system's ability to collect, process and represent knowledge of itself from its subsequent resulting behaviour. We briefly illustrated this in the context of communication strategies for distributed smart cameras in a decentralised network.

The biological background and subsequent working definitions presented in this paper provide steps towards establishing a principled methodology for the consideration of self-awareness and self-expression in the design of autonomous systems. In future work we aim to build upon this by developing a reference architectural framework which will act as a further design tool. In addition, the methodology will provide a rigorous framework for benchmarking of such systems such that different approaches to realising self-awareness and self-expression can be compared with respect to the intended application.

REFERENCES

[1] S. Baker, "The identification of the self." *Psyc. Rev.*, vol. 4, no. 3, pp. 272–284, 1897.

[2] G. A. Tawney, "Feeling and self-awareness." *Psyc. Rev.*, vol. 9, no. 6, pp. 570 – 596, 1902.

[3] A. Morin, "Levels of consciousness and self-awareness : A comparison and integration of various neurocognitive views," *Cons. and Cog.*, vol. 15, no. 2, pp. 358–71, 2006.

[4] S. Duval and R. A. Wicklund, *A theory of objective self awareness*. Academic Press, 1972.

[5] C. Goukens, S. Dewitte, and L. Warlop, "Me, myself, and my choices: The influence of private self-awareness on preference-behavior consistency," Katholieke Universiteit Leuven, Tech. Rep., 2007.

[6] U. Neisser, "The roots of self-knowledge: Perceiving self, it, and thou," *Annals of the NY AoS.*, vol. 818, pp. 19–33, 1997.

[7] A. Morin and J. Everett, "Conscience de soi et langage interieur: Quelques speculations. [self-awareness and inner speech: Some speculations]," *Philosophiques*, vol. XVII, no. 2, pp. 169–188, 1990.

[8] M. Mitchell, "Self-awareness and control in decentralized systems," in *Metacognition in Computation*, 2005, pp. 80–85.

[9] J. Metcalfe and A. P. Shumamura, Eds., *Metacognition: Knowing about knowing*. Cam, MA, USA: MIT Press, 1994.

[10] M. Cox, "Metacognition in computation: A selected research review," *Art. Int.*, vol. 169, no. 2, pp. 104–141, 2005.

[11] L. Paulson, "DARPA creating self-aware computing," *Computer*, vol. 36, no. 3, p. 24, 2003.

[12] E. Amir, M. L. Anderson, and V. K. Chaudhri, "Report on darpa workshop on self-aware computer systems," UIUC Comp. Sci., Tech. Rep. UIUCDCS-R-2007-2810, 2007.

[13] A. Agarwal, J. Miller, J. Eastep, D. Wentziaff, and H. Kasture, "Self-aware computing," MIT, Tech. Rep. AFRL-RI-RS-TR-2009-161, 2009.

[14] A. Agarwal and B. Harrod, "Organic computing," MIT and DARPA, Tech. Rep. White paper, 2006.

[15] L. Esterle, P. R. Lewis, M. Bogdanski, B. Rinner, and X. Yao, "A socio-economic approach to online vision graph generation and handover in distributed smart camera networks," in *13th International Conference on Distributed Smart Cameras (ICDSC 2011)*, Ghent, Belgium, 2011, in press.

[16] B. Fette, *Cognitive radio technology*. Academic Press, 2009.

[17] J. Wang, D. Brady, K. Baclawski, M. Kokar, and L. Lechowicz, "The use of ontologies for the self-awareness of the communication nodes," in *Proceedings of the Software Defined Radio Technical Conference SDR*, vol. 3, 2003.

[18] J. Ye, S. Dobson, and S. McKeever, "Situation identification techniques in pervasive computing: A review," *Pervasive and Mobile Computing*, vol. In Press., 2011.

[19] R. Hamid, S. Maddi, A. Johnson, A. Bobick, I. Essa, and C. Isbell, "A novel sequence representation for unsupervised analysis of human activities," *Artificial Intelligence*, vol. 173, no. 14, pp. 1221 – 1244, 2009.

[20] O. Brdiczka, J. L. Crowley, and P. Reignier, "Learning situation models in a smart home," *IEEE Trans. Sys. Man Cyber. Part B*, vol. 39, pp. 56–63, 2009.

[21] M. Perkowitz, M. Philipose, K. Fishkin, and D. J. Patterson, "Mining models of human activities from the web," in *Proceedings of the 13th international conference on World Wide Web*, 2004, pp. 573–582.

[22] M. Huebscher and J. McCann, "Simulation model for self-adaptive applications in pervasive computing," in *Proceedings of the Database and Expert Systems Applications, 15th International Workshop*. IEEE Computer Society, 2004, pp. 694–698.

[23] F. Mondada, G. C. Pettinaro, A. Guignard, I. W. Kwee, D. Floreano, J.-L. Deneubourg, S. Nolfi, L. M. Gambardella, and M. Dorigo, "Swarm-bot: A new distributed robotic concept," *Autonomous Robots*, vol. 17, pp. 193–221, 2004.

[24] M. Dorigo, V. Trianni, E. Şahin, R. Groß, T. H. Labella, G. Baldassarre, S. Nolfi, J.-L. Deneubourg, F. Mondada, D. Floreano, and L. M. Gambardella, "Evolving self-organizing behaviors for a swarm-bot," *Autonomous Robots*, vol. 17, pp. 223–245, 2004.

[25] S. Kernbach, "From robot swarm to artificial organisms: Self-organization of structures, adaptivity and self-development," in *Symbiotic Multi-Robot Organisms*, P. Levi and S. Kernbach, Eds. Springer, 2010, vol. 7.

[26] E. Vassev and M. Hinchey, "Knowledge representation and awareness in autonomic service-component ensembles – state of the art," in *14th IEEE International Symposium on Object/Component/Service-oriented Real-time Distributed Computing*. IEEE Computer Society, March 2011, pp. 110–119.

[27] H. S. Kim and D. K. Sherman, "Express yourself: Culture and the effect of self-expression on choice," *Journal of Personality and Social Psychology*, vol. 92, no. 1, pp. 1–11, 2007.

[28] S. Chen, C. A. Langner, and R. Mendoza-Denton, "When dispositional and role power fit: implications for self-expression and self-other congruence." *Journal of Personality and Social Psychology*, vol. 96, no. 3, pp. 710–27, 2009.