ESTERLE, L. and BROWN, J.N.A. 2020. I think therefore you are: models for interaction in collectives of self-aware cyber-physical systems. *ACM transactions on cyber-physical systems* [online], 4(4): special issue on self-awareness in resource constrained CPS and regular papers, article 39. Available from: https://doi.org/10.1145/3375403

I think therefore you are: models for interaction in collectives of self-aware cyber-physical systems.

ESTERLE, L. and BROWN, J.N.A.

2020

I Think Therefore You Are: Models for Interaction in Collectives of Self-aware Cyber-physical Systems

LUKAS ESTERLE, Aarhus University, Denmark
JOHN N. A. BROWN, Evolv Technologies, USA

Cyber-physical systems operate in our real world, constantly interacting with the environment and collaborating with other systems. The increasing number of devices will make it infeasible to control each one individually. It will also be infeasible to prepare each of them for every imaginable rapidly unfolding situation. Therefore, we must increase the autonomy of future Cyber-physical Systems. Making these systems self-aware allows them to reason about their own capabilities and their immediate environment. In this article, we extend the idea of the self-awareness of individual systems toward *networked self-awareness*. This gives systems the ability to reason about how they are being affected by the actions and interactions of others within their perceived environment, as well as in the extended environment that is beyond their direct perception. We propose that different levels of networked self-awareness can develop over time in systems as they do in humans. Furthermore, we propose that this could have the same benefits for networks of systems that it has had for communities of humans, increasing performance and adaptability.

CCS Concepts: • **Computer systems organization → Embedded and cyber-physical systems**; *Robotic autonomy*; • **Computing methodologies → Artificial intelligence**; **Distributed artificial intelligence**; • **Human-centered computing** → *Collaborative and social computing*;

Additional Key Words and Phrases: Self-aware systems, networked self-awareness, collaborative systems, autonomous systems, artificial intelligence

ACM Reference format:
Lukas Esterle and John N. A. Brown. 2020. I Think Therefore You Are: Models for Interaction in Collectives of Self-aware Cyber-physical Systems. *ACM Trans. Cyber-Phys. Syst.* 4, 4, Article 39 (June 2020), 25 pages.
https://doi.org/10.1145/3375403

1 INTRODUCING MODELS OF COMPUTATIONAL SELF-AWARENESS

Cyber-physical Systems (CPS) are computational systems specifically designed to interact with their physical environment. The physical environment, however, is subject to unpredictable changes and is also affected by actions and interactions that are beyond the control of the CPS [73, 136]. With the rising number of devices and systems, and the concomitant rise of data, central control will not be feasible [104]. Furthermore, when dealing with rapidly unfolding situations,

This work was supported by the SOLOMON project (Self-Organisation and Learning Online in Mobile Observation Networks) funded by the European Union H2020 Programme under grant agreement no 705020.
Authors' addresses: L. Esterle, Aarhus University, Finlandsgade 22, Aarhus, 8000, Denmark; email: lukas.esterle@eng.au.dk; J. N. A. Brown, Evolv Technologies, 611 Mission St, San Francisco, CA, 94105, USA; email: jnabrown@gmail.com.

2378-962X/2020/06-ART39 $15.00
https://doi.org/10.1145/3375403

a system might not be able to wait for a response from a central controller. Therefore, to ensure that CPS operate continuously, it will be necessary to make them highly autonomous. Introducing computational self-awareness to CPS was proposed as on approach to make them more efficient, resilient, and flexible, and to cope with unforeseen situations [38].

Awareness of self and of others has been widely studied in mammals. It is the ability to identify oneself as an individual within an environment full of peers: equally valuable, equally thoughtful individuals, who may or may not share one's own motivations [28, 44, 94]. Computational self-awareness is the ability of a system to learn about and model its own resources, capabilities, goals, and interactions. Such models can range from being very simplistic, where the system simply keeps track of the current state of different aspects of itself, (e.g., battery level), to complex, multi-dimensional models of itself in its environment, where the system tracks and interrelates feedback of multiple types, from multiple sources [131]. These models could be formal ones (e.g., causal or logical models) or integrate statistical or probabilistic approaches (e.g., Hidden Markov models). The type of model used, depends on the system, its abilities, and the resources it has available. Importantly, systems are expected to exhibit the ability to reason about these models and exploit them to improve their performance. For computing systems, such an ability would increase their usability by enabling them to cope with situations for which they have not been explicitly designed [70, 75, 86, 99]. CPS deployed in the real world interact with their environment and other systems within. Therefore, we propose that introducing an "awareness that, while others may think differently and reach other conclusions, they are still thinking in some way that is as valid as our own" to cyber-physical systems will increase their potential for collaborative effort, further enhancing their ability to adapt to new situations for which none of them had individually been prepared. We argue that such an extension in self-awareness for CPS will allow more autonomous and meaningful interaction with the world and with other systems within it, enabling unanticipated collaborations. We call this new type of computational self-awareness *Networked Self-awareness*.

In humans, self-awareness is not the end of a process. It is one step along the path toward a sense of the self-awareness of others [132]. This process takes place over years in the natural development of children, and can be seen as they begin to understand that the world contains sentient beings, distinct from oneself and from other non-sentient artefacts [96]. This developmental process is discussed in more detail in the next section. An understanding that other sentients are of equal value to ourselves is a natural step in cognitive development that is not always achieved [118]. The resultant impression of uniqueness contributes to the development of delusions of ability, as described in the Dunning-Kruger Effect [71]. This lack of empathy is contributory to narcissistic insensitivity of the feelings (and even the existence) of others [27]. Taken to the extreme, it is one of the key symptoms of psychopathy [16]. Creating CPS, and artificial intelligences (AIs) in general, that are inherently incapable of this logical fallacy should help mitigate the human concerns that machines might rise and overthrow us all, whether maliciously or for our own good.

The current concept of computational self-awareness is heavily based on the work of Neisser [89]. Neisser proposes five levels of self-awareness in humans that, rather than built atop each other, are developed in parallel from earliest infancy. Lewis et al. [76] and Kounev et al. [70] apply these five levels of self-awareness to computational systems. However, current work toward self-awareness in psychology investigates the importance of interactions within a social system [45, 72]. To perform the social interactions that would facilitate the development of self-awareness, each system would benefit from awareness that each other system is also self-aware. In this work we follow and retrace Neisser's, Lewis' and Kounev's steps from a psychological, philosophical and anthropological point of view to identify blind spots that require further investigation. From here, we fill the gaps and propose *networked self-awareness* as an approach for self-aware computing systems to also become sensitive to the existence of others [37].

To date, self-aware computing is being applied in a variety of application areas. For multi-agent systems, several frameworks and architectures that explicitly incorporate self-awareness have been proposed [29, 30, 51, 114]. Self-awareness was also the subject of numerous papers addressing systems- on-chip [32, 65, 113] and multi-core systems [3, 53, 59, 100]. Here, self-awareness is often used as a means to autonomously balance resources, deal with constraints, or schedule tasks. In computing networks, self-awareness is used for two main purposes: to maintain and even increase the quality-of-service and as a means of defense against attacks on the network [1, 46–48, 122]. In pervasive and ubiquitous computing, self-awareness is used to enable individual devices to coordinate with others to achieve common goals [39, 40, 56, 102]. Recently, networked self-awareness has become important to the competitiveness of self-driving cars as their individual sensory data, usually used to develop models about their immediate environment at runtime, is combined into a detailed map of high traffic areas, as in Tesla's "Fleet Learning Network" [10, 105, 106, 135].

However, when systems are deployed in an environment with other, initially unknown, systems—which may or may not have different behaviours and (self-aware) capabilities—each of them will inevitably encounter situations they were not specifically designed to handle. Nevertheless, we expect and desire the system to act upon this newly encountered situation to the best of its abilities. When applying known skills in a known way cannot solve a problem, we propose a novel concept enabling the systems to decompose and reorganise their existing skills in new ways. Iterating through this process is intended to enable the development of new answers to new questions based on prior knowledge.

The remainder of this article is structured as follows. In Section 2, we outline the developmental process of humans and how we explore and discover our environment and establish interactions until we eventually coordinate in teams. In Section 3, we introduce levels of networked self-awareness, enabling future CPS to learn about their environment and how it is affected by other systems within. In Section 4, we propose a novel concept allowing a system to decompose and re-organise its own skills to deal with changes and situations brought about by other systems in the environment. Afterward, we discuss the most important challenges to be tackled to achieve both, networked self-awareness and the required reorganisation of skills in Section 5. We conclude our article in Section 6 with a discussion and a roadmap, pointing out the most important research fields leading toward networked self-aware Cyber-physical Systems.

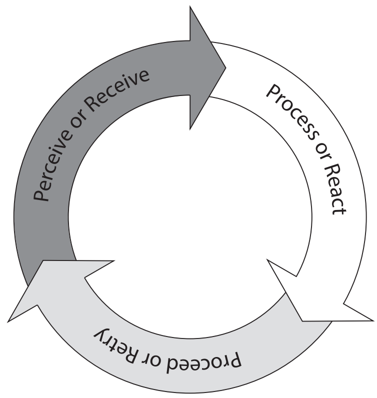
2 DISCOVERING THE WORLD

As the tools of self-awareness develop; as perceptions lead to theories that are tested, proven and disproven in ways that lead to new theories in a rapid and iterative cycle, the human moves naturally from superstition toward logical thinking, and their mental map of the world around them and their place in it changes accordingly. The infant in utero knows nothing but being. Upon birth, the infant experiences fluctuations of cold and hunger. It is only natural to assume that the infant would build a theory of a causal relationship between itself and these sensations in the world: "If I just kick my leg like this and shake my head, and make that screaming sound, then warmth presses against my face, sensation floods into my mouth, and the hunger goes away." This must be an early form of superstition. If I do this and that, then forces beyond my control show mercy and deliver me from hunger. These beliefs eventually succumb as the baby's model of the universe becomes more detailed and accurate through repeated testing. The world has other agents, playing different roles. They can take away hunger or cold. The agent that takes away cold if you pull it over your shoulder and snuggle down into it is not the same agent who takes away cold by picking you up and carrying you across what you will someday call "the room." Perhaps there are multiple agents who pick you up and carry you around. All of them can take away cold and loneliness, and all of them can hug you and press themselves against your face, but only one of them presses your face in the way that takes away hunger. Thus, the categorization of these entities becomes increasingly complex, going from “Me as the World” to “Me and the World” to “Me and the World with others” to “Me and the World with Mommy and Other Agents” and on to “Me and the World with Mommy, and Daddy, and Dog, and other agents in it.” And then the classification changes again, because, while Mommy and Daddy are different. Dog is vastly more different than either of them and a chair even more so. And we only learn later that a chair is inanimate and “responds” differently. We see this in children who attribute volition to the inanimate objects in their lives, until they learn not to [96]. One example is the child who blames an obstacle for tripping them, rather than recognizing that the obstacle has neither motivation nor motive force.

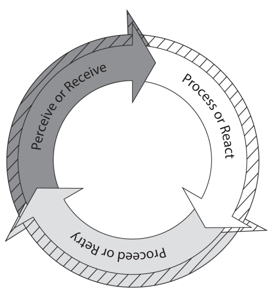
So, the child models interactions with all of these new agents and entities, and her understanding of the world changes as her models are tested again and again over time. This revisiting of previously established models implies the existence and application of a complex memory system of the sort that would allow revision and testing of some memories, separate from and even against all others [96]. While the structures associated with this process have been recognized in humans [54], it is important for the non-biologist to remember that the fundamental neuroplasticity of the human brain has—so far—suggested that direct mapping between brain functions (memory, thought, etc.) and specific anatomical structures is significantly more reliable in computerized systems [101, 111, 121]. We have only a very primitive map of how our brain actually relates to our thoughts, one that has allowed us only to begin the explorations and iterations that will enable us someday to build a complete and useful atlas. In this way, we are fundamentally different from our digital distant cousins. As a result, our models of our own reasoning, and our models of theirs will continue to differ, at least until we learn to quantify the depth, breadth, and width of human thought.

2.1 Time after Time

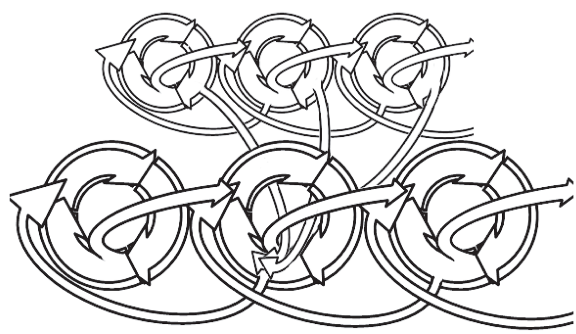
We propose an order of development based on logic rather than observation. Until one perceives that states change over time, one cannot set a goal of influencing or shaping those changes. Let us consider the BRAINS model of human interaction [20] as an illustration of how this might work. The BRAINS model offers a simple way to think of the three fundamentally different speeds at which humans interact with their environment: reflex, reaction, and reflection. Our bodies react reflexively to some environmental change. This all happens in the nervous system, without any conscious or unconscious thought in any part of our brain. For example, a controlled strike to the patellar nerve causes the coordinated relaxation of the femoral biceps and flexion of the femoral quadriceps, making your leg straighten and allowing your doctor to test your reflexes. The same unconscious coordinated reflex can be achieved when a baby’s knees are tickled. Her leg straightens without any thought in any part of the brain. When a baby learns to play with her feet, these reflexive movements combine in more complex patterns, which are triggered in reaction to specific combinations of stimuli. She is learning to recognise and react to complex patterns. She learns to kick in the air and, eventually, to stand and take her first steps. Now the baby is taking conscious control of her reactions, and the straitening movement is reflected upon and deliberately refined as the baby learns to catch her balance and walk. Shortly thereafter, the complex coordination of the muscles involved in walking moves from being under conscious control (requiring deliberate reflection), to being unconscious. The trigger for starting and stopping is deliberate, but the actual coordination has moved into muscle memory, and happens in the hindbrain or cerebellum. This is what allows the child to walk or run without having to concentrate on it. Now that running needs no longer be deliberate, the child can devote her conscious efforts to other goals as she runs. For example, she might add another type of leg-straightening to her routine, and learn to kick a football. With iterative refinement, the deliberate control becomes better and this improvement



(a) A simple feedback loop.



(b) A learning feedback loop, in which one's anticipated cycle of action and reaction (inner loop) is compared to the action/reaction cycle that actually takes place (hatched loop). The understanding that there is a difference is what makes learning possible.



(c) A network of learning feedback loops, in which the stimuli detected by each, and their reactions to those stimuli, all becomes input for the others. Each cell in the network interacts with each other cell, and with the greater environment beyond its individual perception.

Fig. 1. Feedback Loops: Reaction to stimuli can be modelled as a cycle made up of three parts. One or more stimuli are received, the information elicits an internal or external reaction, and the changed system continues to function, ready to react to further stimuli. Figures 1(a) and 1(b) are recreated from Brown [18], Chapter 9. Figure 1(c) from Brown [19].

allows further refinement and so on, in a cycle we call practice. If others are practising at the same time, place, and at a similar or greater level of development, then the child in question can learn to deliberately coordinate her reactions to the actions of others. At first it will be easier to kick a moving ball if she is the only one moving it, but in time, she and her team mates will learn to pass the ball back and forth, anticipating each other's movements in a shared feedback loop. First we perceive change in our immediate environment. Then we react to it. Some of our reactions result in responsive changes to our environment. Once we perceive enough of these effects of our reactions on our environments, iterating and refining the interaction, we learn to deliberately, reflectively modify our reactions to control those effects. The same applies in a team who learn from each other. Through iterative feedback loops, we go from reflex, through reaction, to reflection, and learn to instigate intentional changes to our environment, in either participation or competition with the other creatures in it.

Before we can work with a team, our perception of the world around us has to grow and refine through a natural evolutionary process. Figure 2 illustrates this evolution. The system starts (a) by considering itself as the entire world and everything in it. Soon, it realises that there is more than just itself in the world (b), and that the rest of the world is made up of more than just one entity (c). The fourth step (d) is the beginning of distinguishing between individuals. This can be subdivided into multiple levels of distinction. First, distinguishing living, thinking, and acting objects from inanimate objects. Second, distinguish different types of living creatures (e.g., dog, cat, human) and different types of inanimate objects (e.g., table, chair, toy car I can push). Third, we distinguish different individuals within a type of objects with different granularity (e.g., mum and dad and our dog from the neighbours dog). Finally, we realise that certain, seemingly living/thinking/acting objects are not like humans or dogs (e.g., smart home devices such as Google home or Amazon's Echo that speak to us, or cars and drones that seem to move on their own). Now that we are actively evaluating others, it becomes possible to understand that some of them are evaluating us. We then begin to evaluate their ability to evaluate and compare it to our own. Here, between (d) and (e), some people stop. They remain certain that their ability to think is equal to, or greater than, the thinking ability of all other things. This self-centered delusion is ubiquitous among both children and adults. We are unable to judge our own ignorance of topics about which we are ignorant. This was described by Dunning and Kruger [71], and it has become a common referent in the discussion of social media and recommender systems. It is easier and more comfortable to consider a model of the world in which one's own first-hand experiences and judgements can be allowed to trump the opinions of others who claim to have observed, studied, and measured things beyond our knowledge.

Of course, this requires that we stay in a self-centred world in which the others around us cannot be considered our equals. Those who move forward into awareness that others are also aware and independent from ourselves (e) now live in a world that is much bigger and that contains immeasurably more skills and information than any single individual could possibly acquire in a single lifetime. The result is that our knowledge (red line) expands beyond our immediate perceived environment and the range of potential interactions (dashed red line) (f). We consider this knowledge as *extended environment* while we consider our immediate environment, within the range of potential interactions, as *perceived environment*. Having knowledge of this extended environment and the potential it may offer, allows the self to move into new interactions and learn new things that it will consider to be valuable additions to first-hand knowledge and direct experience.

In Piaget's terms the difference here is between learning that comes from within one's thoughts (Psychogenesis) and learning acquired from outside (Sociogensis). The computational equivalents of these concepts will be discussed in Section 3. For now, let us begin with a simple statement of logic using set-theory to bring some perspective to the value of being able to access information that is generated outside of our own thoughts.

Let us assume $\mathcal{U}_t$ is the set of all perceptible data in the universe at time t. It is safe to assume that this is greater than what we are able to perceive with own senses $\mathcal{I}_t$.

If others exist whose limited senses can perceive information $\mathcal{O}_t$, which lies beyond the range or scope of our abilities, then we can say

$$|\mathcal{I}_t \cap \mathcal{O}_t| \neq 0.$$

Simply speaking, this means one system might be able to perceive information that another system cannot perceive. Range of perception would increase if the systems are themselves in a different

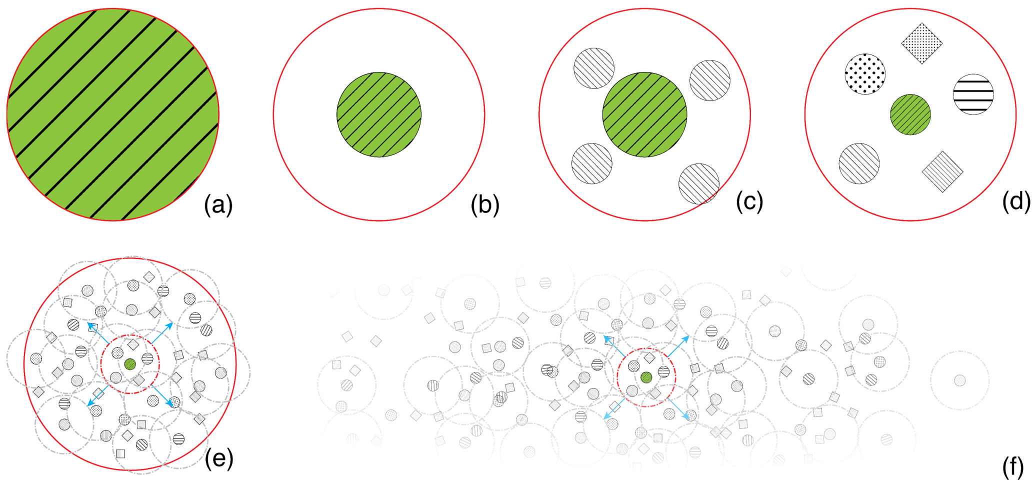


Fig. 2. Developing awareness of the world and of others, from the viewpoint of the self (in green). (a) the self is the entire universe (b) the self realises there is something else, (c) the self becomes aware of others, (d) the self learns that some of these others are distinct from one another through interaction, and (e) the self becomes aware that others are also selves who have their own interactions apart from us. The self has learned that it is not at the centre of the universe, and must now move forward into that much larger, redefined world, or fall backward into the familiar, more comfortable, self-centred perspective. If the self becomes comfortable with and respectful of the awareness of others (f), then it can empathise and learn from their experiences.

locations, or alternatively, if their sensors can perceive different data in the same location. However, if we can share our sensory data, then the amount of shared data must be greater than the individual amount of data perceived

$$|\mathcal{I}_t| < |\mathcal{I}_t \cup \mathcal{O}_t| > |\mathcal{O}_t|.$$

Furthermore, we can turn our sensed data into actionable models and if we can incorporate the sensed information from others, this model becomes a richer and more detailed representation of the world. Further, we can use these models to operate toward our goals. If we are now able coordinate with others and share our tasks to focus our efforts toward a common problem, then we might be able to overcome this problem more quickly.

To apply this concept to real-world CPS, let us return to the example of the autonomous car mentioned earlier. It is understood that a single autonomous car can use streams of sensor data, including ultrasonics, radar, and passive video, to model its immediate environment in real time to mitigate driving hazards such as lane encroachment. If other cars are constructing their models in the same manner, then it is possible for their models of the road-space to be combined in order to cooperate in mitigating hazards that a single vehicle cannot escape. For example, two vehicles driving one behind the other might adjust their speed and relative distance to provide room for a third vehicle in the lane next to theirs to move between them and avoid an incipient crash. With that practical example in mind, let us look briefly at how Psychogenesis and Sociogenesis work together in the healthy human mind to better understand how they could work together in CPS. Being able to compare one's ideas to a larger database of ideas is how humans learn, and it is how humans judge their ability to reason. We can see this in the feedback loops in Figure 1. In the simple version of the feedback loop, one or more stimuli trigger one or more reactions. A more complicated loop follows (Figure 1(b)) in which the individual receives second-hand stimuli from other feedback loops. Now, let us imagine a second layer of each feedback loop (Figure 1(c)) that represents our conscious or unconscious expectations during the cycle. The difference between what we anticipate and what actually happens is the opportunity to learn. If this difference occurs in an internal system, then the resultant learning is psychogenesis. Such is the case when we miss the last step in a staircase and need to quickly adjust our balance to keep from falling. If the difference occurs in interaction between internal and external systems, then the learning is sociogenesis. Such is the case when we miss the last step in a dance and tread on the toes of our partner.

For Cyber-physical Systems to move from a system that is self-centred and does not know about its environment, toward self-aware CPS that also considers other system to have similar capabilities requires research in several different domains. The first step, from a self-centred system (a) to (b) with knowledge of the environment can already be achieved by giving the system knowledge about the environment as well as sensor and simple machine learning capabilities [121]. Returning to automobiles for practical examples of CPS, let us look back in time to illustrate this idea. Early drivers could not tell that their engine was overheating until it did. Adding a simple temperature sensor and a driver-centered output display in which a needle moved from a green zone of safe operating temperature, through yellow and orange zones before reaching a red zone signifying danger, allowed the driver to take action before the engine overheated. Early CPS in automobiles provided combined feedback on engine temperature, oil pressure, and cooling fluid levels by flashing a generic "check engine" message in yellow, orange, or red [57]. The next advance was to enable a live support network of diagnosticians who would be remotely provided with the detailed information that a common driver would not be able to accurately interpret [14]. Systems able to identify other systems and objects in its environment (c) and potentially distinguish them one from another (d) are addressed in research on affordance [21, 64, 126] and in research on life-long machine learning [22, 112, 116, 123]. Moving from the ability to distinguish between objects and other agents (d) to the consideration that other agents are active and have an impact on the environment, is covered in research self-integrating systems [12]. An important aspect of such self-integration is the ability to perform automatic causal reasoning [66, 95] and to determine knowledge of others, e.g., through dynamic epistemic reasoning [129, 130].

A question arises as to whether or not (networked) self-awareness enables and facilitates self-organising and self-optimising behaviour. There is some discussion in the field regarding the differences between networks of systems where each system is capable of reasoning, and networks of unreasoning systems from which spontaneous reasoning may or may not be expected to emerge. In living creatures, this difference has been discussed for millennia. Perhaps the parameters from those discussions can shed some light on the issue of emergence in our field. The ancients debated whether or not fire was alive, because it can be born and, if it does not get enough air or food, it can die. A forest fire seems like a living enemy to those fighting it. It responds to environmental pressures, seeming to sense changes in wind, humidity, atmospheric pressure, and temperature, but this should not be confused with reasoning. As we have recently witnessed in many parts of the world, even the convergence of many fires into the largest conflagrations in recent history has not led to the emergence of intelligence. Similarly, living cells that respond to environmental conditions are not showing rational thought, any more than the exchange of gasses and fluids through 4.5 billion year-old vents between subterranean chambers shows us that the planet is a living thing—despite the fact that these exchanges so closely mirror the microscopic gas exchanges exhibited in the cells of all creatures on the planet and may have been a crucial step in the evolution from geochemistry to biochemistry [124]. We argue that self-organization can take place without self-awareness. Consider the coordinated movements of multiple fields of sunflowers. All around the world, they rotate in synchronicity with the sun. One should not argue that the totality or any of the subsets of flowers is aware of itself, of its peers or of the fact that the radiation that triggers its movement emanates from an astral body. The attribution of complex social and behavioural characteristics to objects or elemental components of living creatures was first formally described by Sir Edward Tylor in 1871 [127]. This aspect of his concept of *Animism* is now often labelled as *Anthropomorphism*, a logical fallacy in which all humans engage at some point in their cognitive development [60]. Related research in computing systems investigates its origins and potential benefits [7].

We believe that a system does not need to be self-aware to self-optimise in a given environment or self-organise with others. However, in quickly unfolding situations, we believe that networked self-awareness will allow systems to achieve better optimisation results when compared to non-self-aware counterparts. Specifically, (networked) self-aware properties, such as understanding the goals and abilities of others and of one's own, can improve the ability of multiple independent systems to self-organise. Those abilities enable systems to find optimal collaborators and to self-organise autonomously during runtime. The benefit of recognising self and others has also been observed in nature [5, 49, 52, 55, 61].

Previous work minimizes or ignores the importance of recognizing that other systems are also self-aware, able to build internal models of themselves, reason upon these models, and hence improve their performance toward their individual goals, irrespectively of the actions, interactions, or goals of others. However, we believe, enabling systems to reason about other systems' self-awareness is key, whether in developing individual self-awareness, cooperating toward common goals, or optimising individual performance to achieve Pareto-efficient outcomes in competing situations.

3 LEVELS OF NETWORKED SELF-AWARENESS

Given the nature of interaction among different devices, entities, and systems, individuals need to be aware of more than just themselves and their immediate environment, they also need an understanding of the impact of their own actions on others and how this reflects back on them. We call this type of self-awareness *Networked Self-awareness* and, similar to the computational self-awareness, as defined by Lewis [75] and Kounev [70], we utilise different levels to categorise a computational system's awareness of its networked environment (cf. Figure 3). In the following we shed light on the different levels and their requirements in terms of understanding their environment as well as the impact that they have on a system. Importantly, the different levels do not have hard and distinct borders. Instead, the characteristics of each level may overlap with others, because old skills are not abandoned, but developed, and modified, and applied in new ways. As an example, networked stimulus awareness gives computational systems the ability to reason about stimuli that impact their environment and, so, have an indirect impact on themselves. This means there is an interaction between at least two elements in the environment. Networked interaction awareness would enable systems to understand that the source of this impact, though unaware of the secondary effects of its actions, is acting with intent and could be induced to cooperate to mutual benefit.

3.1 Networked Stimulus-awareness

Stimulus-awareness allows a system to recognise stimuli impacting itself, regardless of whether they originated internally or externally. *Network stimulus-awareness* allows the system to perceive how stimuli impact the perceived environment, regardless (again) of the internal or external origin of the stimulus. The ability to associate a stimulus to reactions from both the general environment and the individual actors within it is an important step in what psychologists call "calibration." This is the iterative skill of evaluating our own abilities, and it is fundamental to human learning [69]. We might consider different granularities of networked stimulus-awareness. This can originate

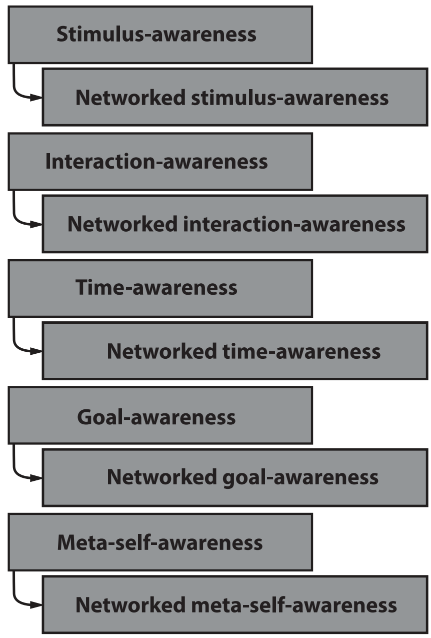


Fig. 3. The different levels of networked self-awareness. There is not necessarily a clear distinction between the different levels and they might sometimes even overlap. Nevertheless, a certain level of computational self-awareness is required to implement the corresponding level of networked self-awareness.

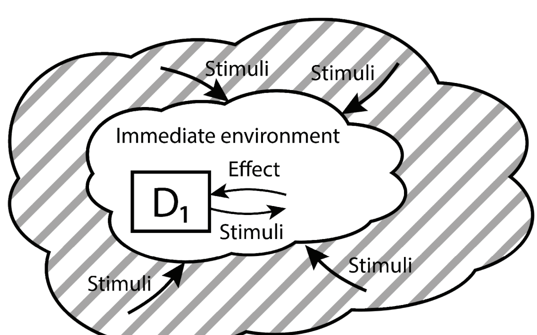


Fig. 4. Networked stimulus awareness enables a system to perceive stimuli in its environment and map them to effects on itself. Where stimuli originate from is unclear at this point but a system knows that these stimuli affect its own immediate known environment.

from being aware of actions and stimuli that affect the perceived environment that have further effects on the system. Networked stimulus-awareness is depicted in Figure 4. We can further refine networked stimulus-awareness, enabling system to understand whether the stimuli has an external or internal origin or whether it originates from within the perceived or the extended environment or even from outside the extended environment. Furthermore, we can improve this such that the system is aware of the source of the stimulus or the recipient within the environment. As discussed in the beginning of this section, knowing both the sources and the action can be considered an awareness of an interaction, which would be part of the networked interaction-awareness.

An example for this would be temperature changes on Systems-on-Chips (SoCs) where the SoC itself might be responsible for changing the temperature but there might also be outside conditions responsible for the temperature change. While at this stage, the SoC might only be able to identify that there is a temperature change, it has no concept of whether this is self-inflicted or caused by an external source [2]. Medical Cyber-physical Systems would be another field where self-awareness has the potential to deliver enormous benefits. Using network stimulus-awareness, the system could reduce hazards through multimodal redundancy, detecting—for example—whether or not an input has been measured correctly through on-board sensors, or if the sensors have detected an accidental or deliberately induced type 1 or type 2 error, as could be the case if the medical CPS were hacked, as was reported in the case of some insulin pumps [134].

3.2 Networked Interaction-awareness

When exhibiting interaction-awareness, a system can reason about its own interactions with the environment including other systems. In *networked interaction-awareness* a system models how the interactions of others affect itself and *vice versa*. This means the system might be able to observe interactions of others, might know the participants of the interactions and, potentially even knows the details and individual actions of their interaction. Importantly, while the interaction might take place even outside the extended environment of a system, it is still aware of the effects the interaction has on its perceived environment or directly on itself.

An early awareness of such interactions is depicted in Figure 5(a). In this first case, D_1 knows that there is an interaction between different entities within its environment that has some kind of affect on it. Figure 5(b) illustrates a system able to trigger this external interaction even though it is unaware of the individual entities performing the interaction. By increasing its awareness of interactions, a system D_1 might potentially even be able to trigger a reaction from D_2. Importantly, this is not just a simple stimulus but active interaction between two or more entities originating from D_2. However, in this early stage of networked interaction-awareness, the system is not aware of all participants of this interaction (cf. Figure 5(c)). This might be due to the fact that some participants of the interaction are outside the perceived environment. Improving this awareness allows the system to trigger this interaction (Figure 5(d)). At a later stage, the system can become aware of a subset or even all participants in the interaction and also how this interaction is performed (cf. Figure 5(e)). Finally, being fully aware of the interactions of others and how they can affect D_1, the system can explicitly trigger the interaction in anticipation of triggering specific reactions and specific effects on its own immediate environment (cf. Figure 5(f)).

One might think of robotic systems able to interact with each other, when two robots interact in a way that aids a third robot, one that is not directly involved in the original interaction. This idea can be continued toward the extent of *social* self-awareness [13] where systems become aware of social concepts such as rules, norms, and traditions of the society they are operating in.

Networked interaction-awareness will be beneficial in many kinds of multi-robot applications, autonomous transportation for goods and passengers, in particular. By exploiting the knowledge gathered through the interaction of other devices, they will be able to plan and revise their own behaviour accordingly [11, 26]. In a similar way, self-organising surveillance systems and multi-robot exploration teams will be able to benefit from networked interaction-awareness [34, 39].

3.3 Networked Time-awareness

Networked time-awareness describes a system that is aware of how everything in its greater environment is susceptible to change over time. Networked time-awareness therefore spans across all other levels of networked self-awareness.

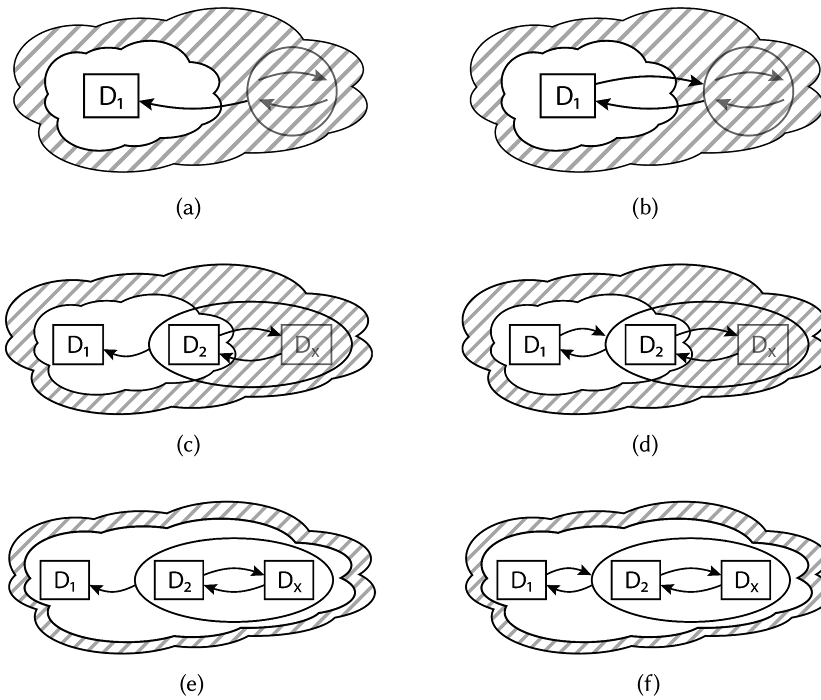


Fig. 5. Different levels of granularity for networked interaction-awareness. Figure 5(a) indicates the system being aware of an interaction even though it does not know about the participants nor the actual actions taking place. Figure 5(b) indicates that the system D_1 is not only aware of the interaction but also knows how to trigger it. Figure 5(c) indicates that D_1 knows about at least one participant of the interaction, and in 5(d) D_1 can trigger the interaction of D_2, Figure 5(e) shows networked interaction awareness where the system D_1 knows about all participants and potentially the performed actions among them that affect it, and Figure 5(f) shows a system D_1 able to trigger an interaction among other systems and is aware of all participants of the interaction.

Figure 6 illustrates the different granularities in this level of networked self-awareness. At first, a system may only be aware of the stimuli happening at different times (6(a)). Refining this information, allows the system to become aware of the temporal order of actions and identify sequence of stimuli within the environment that will eventually have an affect on itself (Figure 6(b)). Further, systems may even become aware of when entire interactions take place that have an effect on the system. This is closely related to networked interaction awareness. However, at first, the system might not be able to distinguish the individual actors and actions of the interactions but only the time T_τ at which the interaction takes place (Figure 6(c)). Figure 6(d) illustrates a system that is aware of interactions, including all their participants and performed actions. Importantly, at this level of granularity, the system is even aware of the sequence of actions performed in the interaction and how they are temporally ordered. This allows the system to have a more fine-grained model of how interactions take place, enabling it to identify potential reasons and sources that affect it. Figure 6(e) shows systems being aware of temporal aspects of goals. A system is aware that goals of others may not arise in parallel at the same time but might be scattered temporally. Having full networked time-awareness, the system is able to identify sequences of goals that affect itself. Having this granularity of awareness also requires a high granularity of networked

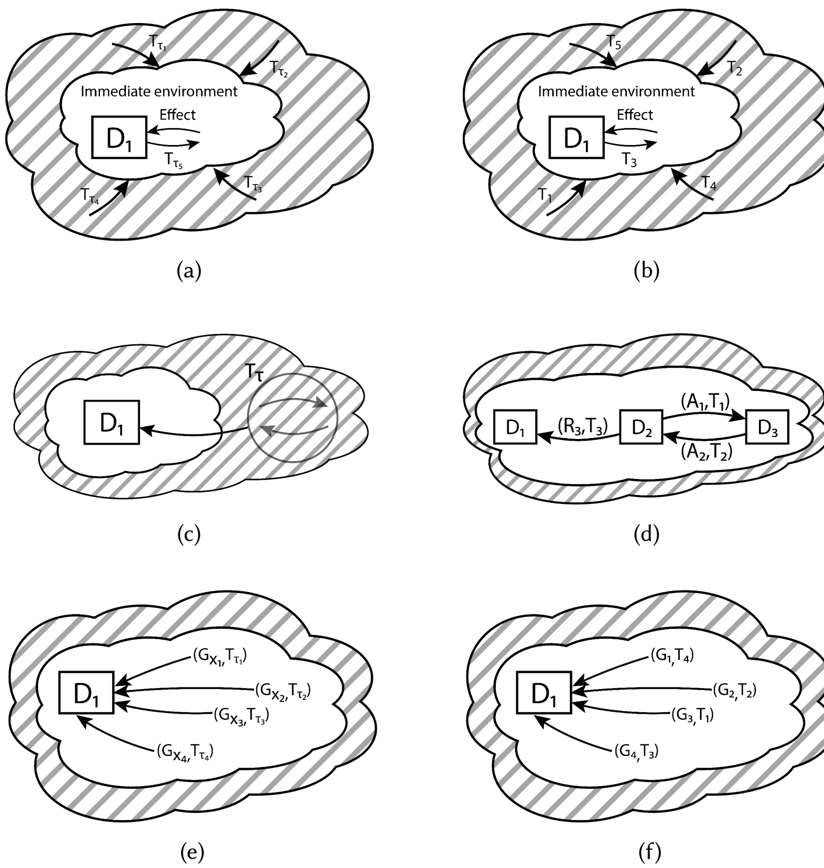


Fig. 6. Different levels of granularity for networked time-awareness. Figure 6(a) shows how a system not only becomes aware of stimuli but also of their different timings τ_i. Importantly, the system is not able to distinguish the order in the sequence of stimuli at this point. Figure 6(b) shows a refined networked time-awareness where a system is able to identify sequences of stimuli as well as their order. Figure 6(c) illustrates the ability to assign timings (T_τ) to general interactions. Only with having refined the ability, the sequence of actions (A) and reactions (R) and their timings within the interaction can be identified (cp. Figure 6(d)). Figure 6(e) illustrates a system being able to assign temporal properties to goals (G_x) and when refining networked time-awareness it will even be aware of sequences of goals ($G_1 - G_4$) within systems as illustrated in Figure 6(f).

goal-awareness. This is illustrated in Figure 6(f). Importantly, a system might even become aware of interdependencies between the different goals due to their temporal properties.

Autonomous production and manufacturing systems, and more specifically autonomous *lean* systems, will require very high flexibility and understanding of time [74, 87, 117]. In a similar way, on autonomous construction sites, where multiple robots build common structures, individual robots need to be able to rely on timely interaction [79]. This will improve overall construction time, keep waiting times to a minimum, and rid the system of faults and collisions.

3.4 Networked Goal-awareness

Goal-awareness allows a system to reason about its own goals. In contrary, networked goal-awareness identifies system able of understanding and knowing the goals of others, how they

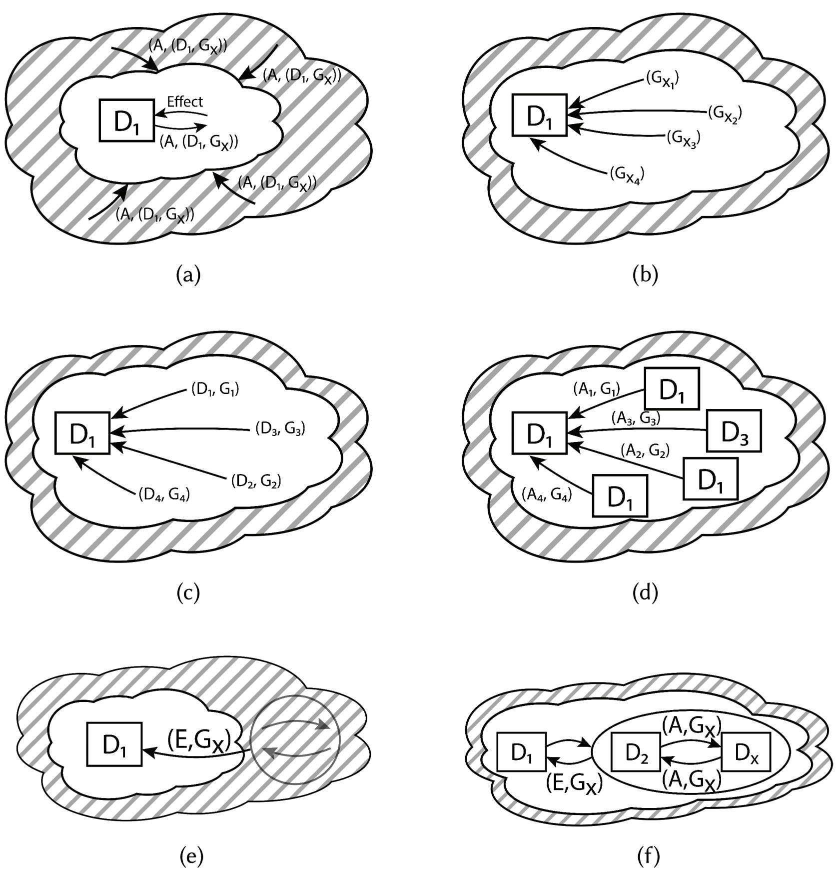


Fig. 7. Different levels of granularity for networked goal-awareness. Figure 7(a) illustrates actions of other, unknown systems have an impact on goals of system D_1. Figure 7(b) shows a system being aware of goals of other systems in its environment. This awareness does not necessarily constitute awareness of the actual systems or how they are exactly mapped to the different goals. The awareness of how systems and goals are mapped is shown in Figure 7(c). Further improving networked goal-awareness, systems can map actions and stimuli from specific systems toward corresponding goals. Finally, systems will be able to map the effects of interactions toward goals. The different granularities are shown in Figure 7(e) and (f).

are pursued, and how this affects them. If a system is aware of the goals of others, then it can start adapting its own actions, behaviour, and even goals to align them to the goals of these other systems. As with the other levels of networked self-awareness, we can conceive different granularities of networked goal-awareness. This may initially be related to own goals and how they are affected by the actions (A in Figure 7) originating from some source in the environment. A system may not be able to distinguish where the stimuli are coming from but knows they affect specific goals (cf. Figure 7(a)). Whether the source of the stimulus is within the perceived or the extended environment or even outside the extended environment might be unclear to the system. This can be due to a lack of knowledge or the ability to generate this knowledge. Figure 7(b) shows a system that is aware of the different goals in the environment. It may initially not know which system correspond to which goal. Only by improving its knowledge of the environment, such a mapping can be achieved (cf. Figure 7(c)) and a system gets aware of the what goals are pursued by which system. Extending its own knowledge about the environment further allows systems to gain awareness how actions performed by other systems to pursue goals affect themselves. This can specifically be problematic when performed actions require shared resources. In a similar way, systems can be aware of the effects (E in Figure 7) on itself, generated by the interactions of others toward their respective goals (cf. Figure 7(e) and (f)).

Having knowledge about the goals of others in the environment and how these other systems pursue those goals affects myself allows for more sophisticated interaction and collaboration among multiple systems. Importantly, devices and systems could improve more than just their own performance. If they are aware of the goals of others, then they can make sure that, so long as their goals do not conflict with each other, their actions do not interfere with one another. If their goals are in conflict, then being aware of this would be the first step toward resolving it. Systems that are aware of the goals of others can develop policies of non-interference, which is an essential step toward benevolent and empathetic self-aware computational systems able to respect the goals of others.

Being aware of the goals of others will become immensely important in future applications for the Internet-of-Things, as well as multi-robot systems such as search-and-rescue or building surveillance [68, 97]. Here, we imagine that individual, autonomous devices collaborate and form teams when they identify the goals they have in common, and the wider range of their disparate abilities [35].

3.5 Networked Meta-self-awareness

Meta-self-awareness is a system able to determine its own level of self-awareness. This ability requires them to have an understanding of the existence of different levels of self-awareness as well as differentiate between them. Employing this ability allows a system to even change its current level of self-awareness. In the same way, we can use networked meta-self-awareness; a system is aware of the existence of the different levels of networked self-awareness. However, we can bring this even one step further and make systems aware of the existence of other systems also being self-aware. Furthermore, system may even be able to distinguish identify and assess the level of self-awareness of another system [36].

While all the above levels of networked self-awareness—as well as self-awareness itself—will be important for future CPS, implementing these features requires a lot of perception, memory, and processing power. Networked meta-self awareness will allow a system to deal with this potential shortcoming. Space systems in particular will benefit from such abilities, since they will be operating under largely unknown circumstances, with limited and shared resources, in environments where manual interactions and adjustments are likely to be conducted by other machines. While we also envision that terrestrial multi-robot systems or applications in the Internet-of-Things will benefit from these abilities, resources are still often likely to be constrained, and direct manual intervention much easier.

4 BUILDING KNOWLEDGE

Our next question, then, should be "How is this knowledge built?" Is it a linear process [82], in which one type of perception is replaced by another type of perception and one mode of thinking is replaced by another mode of thinking, until immature, childish thought is replaced by mature logic, or do types of perception and modes of thinking build up in a cumulative process [108]? Neuroconstructionism [33, 103] theorises that experience changes the structure of the brain. Nativism of Cognitive Architectures [6, 98] is a modification of a theory from evolutionary psychology. Marcus proposes a modification into Neo-Nativism, and suggests the concept of "pre-wiring" rather than "hardwiring" [107]. Farina proposes another alternative, based in part on Roepstorff et al.'s concept of Dynamic Enskilment [109]: Neo-Neuroconstructionism [41]. This is the theory

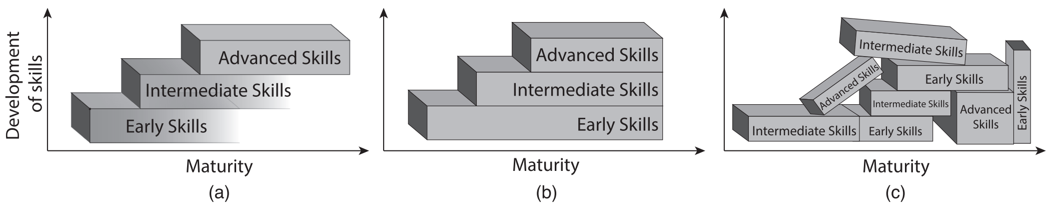


Fig. 8. Three approaches to learn skills. (a) Advanced skills replace previously learned skills, (b) skills build upon each other, and (c) skills can be reused and support each other independent from when they have originally been learnt.

that brain plasticity continues through infancy and adulthood, perhaps using the modular development described by Coltheart [24]. Though some say that can only happen in peripheral processing systems [43], others say it can also happen in the central processing system(s) [6]. We propose that the evolution-based hierarchy of brain structures [18, 80] could allow for the presence of different mechanisms and degrees of plasticity. It seems intuitive that the algorithm that allows your cerebellum to coordinate the complex neuromuscular interactions involved in dribbling a basketball should be implicitly different to the algorithm that allows you to turn this series of letters and spaces into a meaningful phrase. Is there a single, clear barrier between these levels of awareness? If so, then is it a simple binary step, or is it a more gradual transition? Is it a threshold reaction, like the electrochemical exchanges between axon and dendron in our nervous system, or is the transition itself incremental? Once the transition happens, are the old skills replaced, do they change, or is it just that new skills are added? Are cognitive skills more like building blocks that can be stacked one-upon-another so that new heights can be reached (Figure 8(a) and Figure 8(b))?

Consider that a skill is a collection of iterative feedback loops where each individual loop is learned slowly and deliberately and, once well-ingrained, can be run unconsciously as a subroutine. More complex skills are learned as combinations of simpler skills [18]. Consider the example of walking. A child learns the deliberate skill of standing, which consists of finding and maintaining one's balance through a dynamic series of adjustments to keep from falling in any given direction. Once that is learned, the child can learn to walk by keeping from falling in almost every direction. In walking, the child deliberately regains balance during a controlled forward fall by adjusting the placement of first one foot and then the other. When appropriate, simple and complex skills can also be applied in new circumstances with minimal modifications, and without having to be re-learned. Having mastered one skill, one might master a related skill relatively quickly and effortlessly. This would suggest that we can easily transfer certain skills to similar tasks. However, we also speculate that learned skills are made up of new and old cycles of patterned responses; combined like subroutines to accomplish new tasks. Like building blocks, they can be combined and recombined in non-intuitive ways into previously used or completely new patterns, in response to the environment (Figure 8(c)). In a similar way, we would require computational systems to be able to decompose skills, knowledge and capabilities into individual subroutines. To foster collaboration among systems, each individual will also require knowledge, potentially generated at runtime, mapping the skills available from other devices. This will allow systems to efficiently decide who to collaborate with for a given problem. Finally, subroutines need to be combined to establish a new skill online and within the constraints of the current situation.

5 OPEN CHALLENGES

In this position paper we outlined a human-centred approach enabling future CPS to become aware of themselves and of others, based on the latest literature in the field. Introducing *networked self-*

awareness to CPS will increase their autonomy; capitalising on their abilities and resources even when the number of devices and systems will overwhelm human coordinators and centralised controllers. However, there are several challenges that need to be tackled to achieve *self-awareness* and even more to achieve *networked self-awareness*.

One of the biggest challenges to implement and achieve *networked self-awareness* is the limited resources a CPS usually has at its disposal. With limited processing capability and storage on each individual device, observing the environment and generating corresponding models and adapting them at runtime quickly becomes ambitious. Extending this toward different domains (e.g., goals, interactions, or stimuli) makes this task even more complicated.

Another challenge is the inherent uncertainty the CPS face throughout their deployment in the physical world. This uncertainty can manifest itself in different ways:

- **Noise:** *Did the system notice a change in the state of the observed event/situation/object or was this observations caused by environment fluctuation?* Making observations using sensors is usually noisy. Having information about the type of noise can allow modelling and subsequent filtering of the noise. Alternatively, knowing what should be sensed (*ground truth*), the system can identify the noise and filter it in subsequent observations. However, in autonomous, self-aware CPS, the system might not be aware of the sensed information nor of the noise in the environment. The CPS can only gain that awareness by gathering sufficient data over time, and learning to recognise and then filter the noise from the real, physical occurrence of the phenomenon.
- **Timeliness:** *Did the system look long enough, or at the right time?* When attempting to observe a phenomenon, the frequency and duration of sensing becomes crucial in order to make sure it is not missed. If the phenomenon occurs between two measurements, then the system might assume the event or phenomenon did not occur. At the same time, the question arises of how long something should be observed. The system could be very certain that the observed objects, system, or phenomenon behaves in a very specific way after observing it for a longer period of time. Nevertheless, the observed object/system/phenomenon might still change as soon as the CPS stops observing it. (e.g., "I looked at this object for all this time and it did nothing *ergo* it is not self-aware and might actually be just and inanimate object.")
- **Situatedness:** *Does the current location in the real world allow perception of all necessary information?* Being deployed in the real world, each CPS has its very own perception of the environment, including the other systems in that environment. Two systems observing the environment from fundamentally different locations may have a different experience of the same object/system/phenomenon/event. This can easily occur through occlusions and other unanticipated problems, such as a change of noise due to location, and can therefore easily lead to misinterpretation of the situation. (e.g., "That's not what it looked like from my point of view.")
- **Locality:** *Is the system able to perceive and sense enough to draw meaningful conclusions based on events/phenomena?* The principle of locality in physics states that a system can only be directly influenced by its immediate environment. While this view is challenged by quantum physics, it is definitely true for current CPS. However, there it is possible to imagine events that could trigger a chain reaction where the CPS can only perceive the final, immediate event/phenomena. This becomes important when a system can only model the environment (and its events and phenomena) by observing results or intermediate events, without observing the initial actions. (e.g., "I am struck by a toppling stone, unaware that

it is a domino, in a chain of dominoes, put in place and set in motion by players I cannot perceive.")

- **Intention:** *Can a system derive intention behind perceived actions performed by other systems in the environment?* Due to locality, a system might only be able to perceive events and phenomena in its immediate environment. That said, even when a system can observe the root cause of an event or phenomena, it may require further information to be able to decide whether the action was performed intentionally or accidentally (e.g., "The domino that struck me in the previous example had no inherent intent to do so, and we may not be able to derive the intent of the human who pushed the first domino in the chain. Was she made to do it? Did she intend to hit me in this complicated way? Did my being in the path of the final domino actually interfere with her original intent?")
- **Skill Mapping:** *Is a system able to determine with absolute certainty whether specific actions caused a phenomena/event?* Future CPS will be able to learn about their own actions and behaviours. They will also be able to decompose actions and skills and re-compose them to create new behaviours. However, a system deployed in the real world will not be able to decide whether an action it performed resulted in a desired or undesired event, or whether this event was only caused by chance. (e.g., "I combined 2 skills and can now achieve better performance when tackling this task, but which skill did the trick? Or was it something else in the environment of which I am not aware?").

6 ROADMAP AND CONCLUSION

The list presented in the previous section only represents the most important challenges that need to be tackled to achieve initial *networked self-awareness*. Some of these challenges are being tackled by researchers either directly or partially. For example, systems could rely on the information of others to determine the level of self-awareness of another system or to mitigate *Situatedness* or *Locality*. However, relying on the information of others raises the issue of trust in the information-providing systems. To tackle this issue, mechanisms for computational trust are applied [4].

In the following, we outline a roadmap of research areas required to achieve the initial steps toward *networked self-awareness* for Cyber-physical Systems:

- **Explainable AI:** Initially conceived as an approach for machines to explain and rationalise their behaviour to humans; having systems with an ability to explain their own behaviour allows humans to understand the decision and reasoning processes of the machine [58, 133]. Work in this area will also allow for machines having the ability to explain themselves to other machines. However, machines will require common semantics. If they do not have such a semantics, then they require the ability and possibility to generate such a common understanding themselves.
- **Runtime modelling:** CPS need to be able to model various aspects of themselves and their environment at runtime. These models will need to include ongoing internal and external readings [15]. As a next step they will need to model their own behaviour, and that of others in the environment [17, 31, 128]. Furthermore, they will need to generate causal models to track the reasons for certain behaviours and interactions [66]. While there is a rich body of research around modelling various aspects of CPS at runtime, modelling one's own goals, as well as the goals of others seems to be a largely unexplored research area.
- **Sensor and model fusion:** There has been a lot of research on fusing information from numerous sensors, whether on single systems or distributed among multiple systems [92, 110, 119]. However, this assumes that the syntax and semantics of diverse sensors are understood across the different devices. In a similar way, CPS have to be able to agree on

models before they can fuse them. There remains a risk of misunderstanding, and a risk that models might be incorrectly fused. There would be a requirement that mechanisms could decide whether or not to rely on fused information. This could be the case when a CPS is in a non-critical state and, the use of a potentially incorrect model could not harm either the system, or anything or anyone in the environment. This would, again and in turn, rely upon well-established models of the environment.

- **Affordance:** The ability to learn about the current environment and how it can be used toward one's own goals is called affordance [50]. In robotic systems there is a large research community working on the problem of learning about the environment and how to exploit it to more quickly achieve one's own goals [64, 77, 85, 88]. Importantly, these goals are all predefined, and each of the robots is operating toward their own specific, individual goals.
- **Runtime goal generation, decomposition, and adaptation:** As with current computing, Cyber-physical Systems are designed to operate toward specific goals. Systems usually become idle as soon as these goals are satisfied, and remain so until the goals are either no longer satisfied or replaced. We argue that, once future systems have reached their initial goals, they will be required to generate their own further goals during runtime. This needs to be done in a way that enables them to remain available to the stakeholder. In addition, we argue that, when cooperating with other systems, each individual will need to be able to decompose and adapt its own goals. This will become specifically important when systems have conflicting goals and run into stalemates or deadlocks.
- **Runtime state estimation and verification:** Having systems estimate their current state, and verify this estimate during runtime, is a core topic of CPS [9, 23, 62, 67, 90, 120]. This has been extended toward collaborative verification and state estimation in distributed CPS. Estimating and verifying state requires a definition of how the system should behave [8]. Generating these definitions during runtime for individual systems embedded in a highly dynamic environment requires an understanding of the situatedness and location of the systems. Establishing this at runtime remains a difficult and unsolved problem.
- **Knowledge de-composition and re-organisation:** In transfer learning, the goal is to apply previously learned knowledge in new tasks. This has the benefit of getting a better initial performance and learned models only need to be adapted to new situations [25, 93, 125]. However, this approach assumes that the new task fits to the previously learned knowledge at least partially. In other cases, where the learned knowledge does not fit at all with the new task, this can have the opposite effect. An alternative approach to this would be to de-compose tasks into smaller chunks and learn models for each individual task. In case of new goals arising where different tasks have to be tackled, previously de-composed tasks can be recompiled to achieve this goal. Autonomous task-decomposition is a first step in this direction [63].
- **Online plan synthesis:** In dynamically changing environments, next actions should be carefully planned. Online plan synthesis for multi-robot systems in uncertain environments has only recently received a lot of interest [42, 78, 91, 115]. However, this still assumes knowledge about the environment and the available robots or relies on central controller and planner. These assumptions have only been relaxed recently [83, 84] and need further relaxation to achieve networked self-aware CPS.
- **Computational trust:** Trust among machines is tackled with mechanisms of computational trust [4, 81]. Future, self-aware machines will be required to reason about these trust mechanisms and apply them differently depending on the machine they are expected to trust. Another question arising is how to deal with machines deliberately but falsely trying to generate trust in themselves.

This roadmap summarises the areas currently under research; areas intended to lead, eventually, to networked self-aware Cyber-physical systems. Of course, this list will be extended over time as new research fields arise and novel challenges are identified. In this article, we have presented two novel ideas, namely networked self-awareness and knowledge de-composition and re-organisation. We discussed different levels of networked self-awareness inspired by both the original levels described by Neisser [89] and the computational levels described by Lewis [75]. We identified the remaining challenges to the achievement of networked self-aware CPS, and outlined a roadmap to overcome them. Self-awareness and networked self-awareness will make future CPS more resilient, more robust, and more dependable by allowing them to develop a deeper understanding of the dynamic environments in which they will operate, and the quickly unfolding situations with which they will be faced. These techniques will also enable systems to self-integrate during runtime, mitigating extensive deployment efforts and human intervention [104]. In those situations, future CPS will need to balance the pursuit of their primary goals against the pursuit of improved performance through the development of a better understanding of their environment and how to interact with it. In this way, Networked Self-aware Systems will be able to weigh their own options and decide whether or not to spend resources on the latter in order to improve their ability to do the former.

REFERENCES

[1] Sani Abba and Jeong-A. Lee. 2015. An autonomous self-aware and adaptive fault tolerant routing technique for wireless sensor networks. *Sensors* 15, 8 (2015), 20316–20354. DOI: https://doi.org/10.3390/s150820316

[2] Andreas Agne, Markus Happe, Achim Löosch, Christian Plessl, and Marco Platzner. 2016. *Self-aware Compute Nodes*. Springer International Publishing, Cham, 145–165. DOI: https://doi.org/10.1007/978-3-319-39675-0_8

[3] Andreas Agne, Markus Happe, Achim Lösch, Christian Plessl, and Marco Platzner. 2014. Self-awareness as a model for designing and operating heterogeneous multicores. *ACM Trans. Reconfig. Technol. Syst.* 7, 2, Article 13 (2014), 18 pages. DOI: https://doi.org/10.1145/2617596

[4] Peter Andras, Lukas Esterle, Michael Guckert, The Anh Han, Peter R. Lewis, Kristina Milanovic, Terry Payne, Cedric Perret, Jeremy Pitt, Simon T. Powers, Neil Urquhart, and Simon Wells. 2018. Trusting intelligent machines: Deepening trust within socio-technical systems. *IEEE Technology and Society Magazine* 37, 4 (2018), 76–83.

[5] Valeriu Balaban, Sean Lim, Gaurav Gupta, James Boedicker, and Paul Bogdan. 2018. Quantifying emergence and self-organisation of Enterobacter cloacae microbial communities. *Sci. Rep.* 8, 12416 (2018), 1–9.

[6] Jerome H. Barkow, Leda Cosmides, and John Tooby. 1995. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press, New York, NY.

[7] Chloe M. Barnes, Lukas Esterle, and John N. A. Brown. 2019. When you believe in things that you don't understand: The effect of cross-generational habits on self-improving system integration. In *Proceedings of the IEEE International Workshops on Foundations and Applications of Self* Systems (FAS*W'19)*. 1–4.

[8] Ezio Bartocci, Jyotirmoy Deshmukh, Alexandre Donzé, Georgios Fainekos, Oded Maler, Dejan Ničković, and Sriram Sankaranarayanan. 2018. *Specification-Based Monitoring of Cyber-Physical Systems: A Survey on Theory, Tools and Applications*. Springer International Publishing, Cham, 135–175. DOI: https://doi.org/10.1007/978-3-319-75632-5_5

[9] Ezio Bartocci, Radu Grosu, Atul Karmarkar, Scott A. Smolka, Scott D. Stoller, Erez Zadok, and Justin Seyster. 2013. Adaptive runtime verification. In *Runtime Verification*, Shaz Qadeer and Serdar Tasiran (Eds.). Springer, Berlin, 168–182.

[10] Mohamad Baydoun, Mahdyar Ravanbakhsh, Damian Campo, Pablo Marin, David Martin, Lucio Marcenaro, Andrea Cavallaro, and Carlo S. Regazzoni. 2018. A multi-perspective approach to anomaly detection for self-aware embodied agents. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'18)*. IEEE, 6598–6602.

[11] Basil Becker and Holger Giese. 2008. On safe service-oriented real-time coordination for autonomous vehicles. In *Proceedings of the 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing*. 203–210. DOI: https://doi.org/10.1109/ISORC.2008.13

[12] Kirsti L. Bellman, Jean Botev, Ada Diaconescu, Lukas Esterle, Christian Gruhl, Christopher Landauer, Peter R. Lewis, Anthony Stein, Sven Tomforde, and Rolf P. Würtz. 2018. Self-improving system integration—Status and challenges after five years of SISSY. In *Proceedings of the IEEE International Conference on Self-organizing and Self-Adaptive Systems Workshop*. 160–167. DOI: https://doi.org/10.1109/FAS-W.2018.00042

[13] Kirstie L. Bellman, Jean Botev, Hanno Hildmann, Peter R. Lewis, Steven Marsh, Jeremy Pitt, Ingo Scholtes, and Sven Tomforde. 2017. Socially-sensitive systems design: Exploring social potential. *IEEE Technol. Soc. Mag.* 36, 3 (2017), 72–80.

[14] Ivan Berger. 2002. Can you trust your car? [automotive electronics reliability testing]. *IEEE Spectr.* 39, 4 (2002), 40–45. DOI: https://doi.org/10.1109/6.993787

[15] Gordon Blair, Nelly Bencomo, and Robert B. France. 2009. Models@ run.time. *Computer* 42, 10 (2009), 22–27. DOI: https://doi.org/10.1109/MC.2009.326

[16] Robert James and Robert J. R. Blair. 1995. A cognitive developmental approach to morality: Investigating the psychopath. *Cognition* 57, 1 (1995), 1–29.

[17] Darko Bozhinoski, David Garlan, Ivano Malavolta, and Patrizio Pelliccione. 2019. Managing safety and mission completion via collective run-time adaptation. *J. Syst. Arch.* 95 (2019), 19–35. DOI: https://doi.org/10.1016/j.sysarc.2019.02.018

[18] John N. A. Brown. 2015. *Anthropology-based Computing. Springer, Cham.*

[19] John N. A. Brown. 2015. Making sense of the noise: An ABC approach to big data and security. In *Application of Big Data for National Security*, Babak Akhgar, Gregory B. Saathoff, Hamid R. Arabnia, Richard Hill, Andrew Staniforth, and Petra Saskia Bayerl (Eds.). Butterworth-Heinemann, 261–273. DOI: https://doi.org/10.1016/B978-0-12-801967-2.00018-5

[20] John N. A. Brown. 2016. *Psychology and Neurology: The Surprisingly Simple Science of Using Your Brain.* Springer International Publishing, Cham, 103–118.

[21] Chenyi Chen, Ari Seff, Alain Kornhauser, and Jianxiong Xiao. 2015. DeepDriving: Learning affordance for direct perception in autonomous driving. In *Proceedings of the IEEE International Conference on Computer Vision.*

[22] Zhiyuan Chen and Bing Liu. 2016. Lifelong machine learning. *Synth. Lect. Artif. Intell. Mach. Learn.* 10, 3 (2016), 1–145. DOI: https://doi.org/10.2200/S00737ED1V01Y201610AIM033

[23] Matthew Clark, Xenofon Koutsoukos, Joseph Porter, Ratnesh Kumar, George Pappas, Oleg Sokolsky, Insup Lee, and Lee Pike. 2013. *A Study on Run Time Assurance for Complex Cyber Physical Systems.* Technical Report.

[24] Max Coltheart. 1999. Modularity and cognition. *Trends Cogn. Sci.* 3, 3 (1999), 115–120.

[25] Wenyuan Dai, Ou Jin, Gui-Rong Xue, Qiang Yang, and Yong Yu. 2009. EigenTransfer: A unified framework for transfer learning. In *Proceedings of the 26th Annual International Conference on Machine Learning.* ACM, 193–200. DOI: https://doi.org/10.1145/1553374.1553399

[26] Wilhelm Dangelmaier, Holger Giese, Florian Klein, Hendrik Renken, and Peter Scheideler. 2004. Shared experiences in intelligent transportation systems. *IFAC Proceedings* 37, 8 (2004), 142–147. DOI: https://doi.org/10.1016/S1474-6670(17)31965-1

[27] Jean Decety and Philip L Jackson. 2004. The functional architecture of human empathy. *Behav. Cogn. Neurosci. Rev.* 3, 2 (2004), 71–100.

[28] David DeGrazia. 2009. Self-awareness in animals. In *The Philosophy of Animal Minds.* Cambridge University Press.

[29] Ada Diaconescu, Kirstie L Bellman, Lukas Esterle, Holger Giese, Sebastian Götz, Peter Lewis, and Andrea Zisman. 2017. *Architectures for Collective Self-aware Computing Systems.* Springer International Publishing, Cham, 191–235. DOI: https://doi.org/10.1007/978-3-319-47474-8_7

[30] Ada Diaconescu, Berry Porter, Roberto Rodrigues, and Evangelos Pournaras. 2018. Hierarchical self-awareness and authority for scalable self-integrating systems. In *Proceedings of the IEEE 3rd International Workshops on Foundations and Applications of Self* Systems (FAS*W'18).* 168–175. DOI: https://doi.org/10.1109/FAS-W.2018.00043

[31] Nicolás R. D'Ippolito, Victor Braberman, Nir Piterman, and Sebastián Uchitel. 2010. Synthesis of live behaviour models. In *Proceedings of the 18th ACM SIGSOFT International Symposium on Foundations of Software Engineering.* ACM, New York, NY, 77–86.

[32] Nikil Dutt, Axel Jantsch, and Santanu Sarma. 2016. Toward smart embedded systems: A self-aware system-on-chip (SoC) perspective. *ACM Trans. Embed. Comput. Systems* 15, 2, Article 22 (2016), 27 pages. DOI: https://doi.org/10.1145/2872936

[33] Jeffrey L. Elman, Elizabeth A. Bates, and Mark H. Johnson. 1998. *Rethinking Innateness: A Connectionist Perspective on Development.* Vol. 10. MIT Press, Cambridge, MA.

[34] Lukas Esterle. 2017. Centralised, decentralised, and self-organised coverage maximisation in smart camera networks. In *Proceedings of the IEEE 11th International Conference on Self-Adaptive and Self-Organizing Systems.* 1–10. DOI: https://doi.org/10.1109/SASO.2017.9

[35] Lukas Esterle. 2018. Goal-aware team affiliation in collectives of autonomous robots. In *Proceedings of the IEEE 12th International Conference on Self-Adaptive and Self-Organizing Systems.* 90–99. DOI: https://doi.org/10.1109/SASO.2018.00020

[36] Lukas Esterle, Kirstie L. Bellman, Steffen Becker, Anne Koziolek, Christopher Landauer, and Peter Lewis. 2017. *Assessing Self-awareness.* Springer International Publishing, Cham, 465–481.

[37] Lukas Esterle and John N. A. Brown. 2018. Levels of networked self-awareness. In *Proceedings of the IEEE International Workshops on Foundations and Applications of Self* Systems (FAS*W'18)*. 237–238.

[38] Lukas Esterle and Radu Grosu. 2016. Cyber-physical systems: Challenge of the 21st century. *e & i Elektrotech. Informationstech.* 133, 7 (2016), 299–303.

[39] Lukas Esterle and Bernhard Rinner. 2018. An architecture for self -aware IOT applications. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. 6588–6592. DOI: https://doi.org/10.1109/ICASSP.2018.8462005

[40] Lukas Esterle, Jennifer Simonjan, Georg Nebehay, Roman Pflugfelder, Gustavo Fernández Domínguez, and Bernhard Rinner. 2016. *Self-aware Object Tracking in Multi-Camera Networks*. Springer International Publishing, Cham, 261–277. DOI: https://doi.org/10.1007/978-3-319-39675-0_13

[41] Mirko Farina. 2014. Three approaches to human cognitive development: Neo-nativism, neuroconstructivism, and dynamic enskillment. *Br. J. Philos. Sci.* 67, 2 (2014), 617–641.

[42] Alexandros Filotheou, Alexandros Nikou, and Dimos V. Dimarogonas. 2018. Decentralized control of uncertain multi-agent systems with connectivity maintenance and collision avoidance. In *European Control Conference (ECC'18)*. IEEE, 8–13.

[43] Jerry A. Fodor. 2001. *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. MIT Press, Cambridge, MA.

[44] Gordon G. Gallup. 1982. Self-awareness and the emergence of mind in primates. *Am. J. Primatol.* 2, 3 (1982), 237–248.

[45] Gordon G. Gallup. 1998. Self-awareness and the evolution of social intelligence. *Behav. Process.* 42, 2–3 (1998), 239–247.

[46] Erol Gelenbe. 2009. Steps toward self-aware networks. *Commun. ACM* 52, 7 (2009), 66–75. DOI: https://doi.org/10.1145/1538788.1538809

[47] Erol Gelenbe, Michael Gellman, and Pu Su. 2003. Self-awareness and adaptivity for quality of service. In *Proceedings of the 8th IEEE Symposium on Computers and Communications*. Vol. 1. DOI: https://doi.org/10.1109/ISCC.2003.1214093

[48] Erol Gelenbe, Ricardo Lent, and Arturo Nunez. 2004. Self-aware networks and QoS. *Proc. IEEE* 92, 9 (2004), 1478–1489. DOI: https://doi.org/10.1109/JPROC.2004.832952

[49] Karine A. Gibbs, Mark L. Urbanowski, and E. Peter Greenberg. 2008. Genetic determinants of self identity and social recognition in bacteria. *Science* 321, 5886 (2008), 256–259. DOI: https://doi.org/10.1126/science.1160033

[50] James J. Gibson. 1966. The senses considered as perceptual systems. Greenwood Press.

[51] Holger Giese, Thomas Vogel, Ada Diaconescu, Sebastian Götz, and Samuel Kounev. 2017. *Architectural Concepts for Self-aware Computing Systems*. Springer International Publishing, Cham, 109–147. DOI: https://doi.org/10.1007/978-3-319-47474-8_5

[52] Martin Giurfa, Shaowu Zhang, Arnim Jenett, Randolf Menzel, and Mandyam V. Srinivasan. 2001. The concepts of 'sameness' and 'difference' in an insect. *Nature* 410, 6831 (2001), 930–933.

[53] Maximilian Götzinger, Amir M. Rahmani, Martin Pongratz, Pasi Liljeberg, Axel Jantsch, and Hannu Tenhunen. 2016. The role of self-awareness and hierarchical agents in resource management for many-core systems. In *Proceedings of the IEEE 10th International Symposium on Embedded Multicore/Many-core Systems-on-Chip*. 53–60. DOI: https://doi.org/10.1109/MCSoC.2016.57

[54] Ann M. Graybiel. 1995. Building action repertoires: Memory and learning functions of the basal ganglia. *Curr. Opin. Neurobiol.* 5, 6 (1995), 733–741. DOI: https://doi.org/10.1016/0959-4388(95)80100-6

[55] Hans J. Gross, Mario Pahl, Aung Si, Hong Zhu, Jürgen Tautz, and Shaowu Zhang. 2009. Number-based visual generalisation in the honeybee. *PLoS ONE* 4, 1 (01 2009), 1–9. DOI: https://doi.org/10.1371/journal.pone.0004263

[56] Guettatfi Guettatfi, Philipp Hübner, Marco Platzner, and Bernhard Rinner. 2017. Computational self-awareness as design approach for visual sensor nodes. In *Proceedings of the 12th International Symposium on Reconfigurable Communication-centric Systems-on-Chip*. 1–8. DOI: https://doi.org/10.1109/ReCoSoC.2017.8016147

[57] James J. Gumbleton and Lauren L. Bowler. 1982. General motors' computer command control-system development. In *SAE Technical Paper*. DOI: https://doi.org/10.4271/820901

[58] David Gunning. 2017. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), nd Web* (2017).

[59] Markus Happe, Andreas Agne, Christian Plessl, and Marco Platzner. 2012. Hardware/software platform for self-aware compute nodes. In *Proceedings of the Workshop on Self-Awareness in Reconfigurable Computing Systems (SRCS'12)*. 8.

[60] Fritz Heider and Marianne Simmel. 1944. An experimental study of apparent behavior. *Am. J. Psychol.* 57, 2 (1944), 243–259.

[61] Bernd Heinrich and Thomas Bugnyar. 2007. Just how smart are ravens? *Sci. Am.* 296, 4 (2007), 64–71.

[62] Radoslav Ivanov, James Weimer, Rajeev Alur, George J. Pappas, and Insup Lee. 2018. Verisig: Verifying safety properties of hybrid systems with neural network controllers. *CoRR* abs/1811.01828 (2018).

[63] Robert A. Jacobs, Michael I. Jordan, and Andrew G. Barto. 1991. Task decomposition through competition in a modular connectionist architecture: The what and where vision tasks. *Cogn. Sci.* 15, 2 (1991), 219–250.

[64] Lorenzo Jamone, Emre Ugur, Angelo Cangelosi, Luciano Fadiga, Alexandre Bernardino, Justus Piater, and José Santos-Victor. 2018. Affordances in psychology, neuroscience, and robotics: A survey. *IEEE Trans. Cogn. Dev. Syst.* 10, 1 (2018), 4–25. DOI: https://doi.org/10.1109/TCDS.2016.2594134

[65] Axel Jantsch, Nikil Dutt, and Amir M. Rahmani. 2017. Self-awareness in systems on chip: A survey. *IEEE Des. Test* 34, 6 (2017), 8–26. DOI: https://doi.org/10.1109/MDAT.2017.2757143

[66] Severin Kacianka and Alexander Pretschner. 2018. Understanding and formalizing accountability for cyber-physical systems. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics.* 3165–3170. DOI: https://doi.org/10.1109/SMC.2018.00536

[67] Kenan Kalajdzic, Ezio Bartocci, Scott A. Smolka, Scott D. Stoller, and Radu Grosu. 2013. Runtime verification with particle filtering. In *Runtime Verification*, Axel Legay and Saddek Bensalem (Eds.). Springer, Berlin, 149–166.

[68] Asif Khan, Bernhard Rinner, and Andrea Cavallaro. 2018. Cooperative robots to observe moving targets: Review. *IEEE Trans. Cybernet.* 48, 1 (2018), 187–198. DOI: https://doi.org/10.1109/TCYB.2016.2628161

[69] Asher Koriat, Sarah Lichtenstein, and Baruch Fischhoff. 1980. Reasons for confidence. *J. Exp. Psychol.: Hum. Learn. Mem.* 6, 2 (1980), 107.

[70] Samuel Kounev, Peter Lewis, Kirstie L. Bellman, Nelly Bencomo, Javier Camara, Ada Diaconescu, Lukas Esterle, Kurt Geihs, Holger Giese, Sebastian Götz, et al. 2017. The notion of self-aware computing. In *Self-Aware Computing Systems.* Springer, Berlin, 3–16.

[71] Justin Kruger and David Dunning. 1999. Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J. Pers. Soc. Psychol.* 77, 6 (1999), 1121.

[72] Mark R. Leary and Nicole R. Buttermore. 2003. The evolution of the human self: Tracing the natural history of self-awareness. *J. Theory Soc. Behav.* 33, 4 (2003), 365–404.

[73] Edward A. Lee. 2008. Cyber physical systems: Design challenges. In *Proceedings of the International Symposium on Object and Component-Oriented Real-Time Distributed Computing*, Vol. 00. 363–369.

[74] Jay Lee, Behrad Bagheri, and Hung-An Kao. 2015. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manufacturing Letters* 3, 2015 (2015), 18–23. DOI: https://doi.org/10.1016/j.mfglet.2014.12.001

[75] Peter R. Lewis, Arjun Chandra, Shaun Parsons, Edward Robinson, Kyrre Glette, Rami Bahsoon, Jim Tørresen, and Xin Yao. 2011. A survey of self-awareness and its application in computing systems. In *Proceedings of the International Conference on Self-Adaptive and Self-Organizing Systems Workshops.* 102–107.

[76] Peter R. Lewis, Marco Platzner, Bernhard Rinner, Jim Tørresen, and Xin Yao (Eds.). 2016. *Self-aware Computing Systems: An Engineering Approach.* Springer.

[77] Manuel Lopes, Francisco S. Melo, and Luis Montesano. 2007. Affordance-based imitation learning in robots. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems.* 1015–1021. DOI: https://doi.org/10.1109/IROS.2007.4399517

[78] Anna Lukina, Lukas Esterle, Christian Hirsch, Ezio Bartocci, Junxing Yang, Ashish Tiwari, Scott A. Smolka, and Radu Grosu. 2017. ARES: Adaptive receding-horizon synthesis of optimal plans. In *Tools and Algorithms for the Construction and Analysis of Systems*, Axel Legay and Tiziana Margaria (Eds.). Springer, Berlin, 286–302.

[79] Alan M. Lytle and Kamel S. Saidi. 2007. NIST research in autonomous construction. *Autonom. Robots* 22, 3 (01 Apr. 2007), 211–221. DOI: https://doi.org/10.1007/s10514-006-9003-x

[80] Paul D. MacLean. 1985. Evolutionary psychiatry and the triune brain. *Psychol. Med.* 15, 2 (1985), 219–221.

[81] Stephen Paul Marsh. 1994. *Formalising Trust as a Computational Concept.* PhD Thesis.

[82] George H. Mead. 1934. *Mind, Self, and Society: From the Standpoint of a Social Behaviorist (Works of George Herbert Mead, Vol. 1).*

[83] Claudio Menghi, Sergio Garcia, Patrizio Pelliccione, and Jana Tumova. 2018. Multi-robot LTL planning under uncertainty. In *Formal Methods*, Klaus Havelund, Jan Peleska, Bill Roscoe, and Erik de Vink (Eds.). Springer International Publishing, Cham, 399–417.

[84] Claudio Menghi, Sergio García, Patrizio Pelliccione, and Jana Tumova. 2018. Towards multi-robot applications planning under uncertainty. In *Proceedings of the 40th International Conference on Software Engineering.* ACM, New York, NY, 438–439. DOI: https://doi.org/10.1145/3183440.3195046

[85] Huaqing Min, Chang'an Yi, Ronghua Luo, Jinhui Zhu, and Sheng Bi. 2016. Affordance research in developmental robotics: A survey. *IEEE Trans. Cogn. Dev. Syst.* 8, 4 (2016), 237–255. DOI: https://doi.org/10.1109/TCDS.2016.2614992

[86] Melanie Mitchell. 2005. Self-awareness and control in decentralized systems. In *Proceedings of the AAAI Spring Symposium: Metacognition in Computation.* 80–85.

[87] László Monostori. 2014. Cyber-physical production systems: Roots, expectations and R&D challenges. *Procedia CIRP* 17 (2014), 9–13. DOI: https://doi.org/10.1016/j.procir.2014.03.115

[88] Luis Montesano, Manuel Lopes, Alexandre Bernardino, and José Santos-Victor. 2008. Learning object affordances: From sensory–motor coordination to imitation. *IEEE Trans. Robot.* 24, 1 (2008), 15–26. DOI: https://doi.org/10.1109/TRO.2007.914848

[89] Ulric Neisser. 1988. Five kinds of self-knowledge. *Philos. Psychol.* 1, 1 (1988), 35–59.

[90] Laura Nenzi, Simone Silvetti, Ezio Bartocci, and Luca Bortolussi. 2018. A robust genetic algorithm for learning temporal specifications from data. In *Quantitative Evaluation of Systems*, Annabelle McIver and Andras Horvath (Eds.). Springer International Publishing, 323–338.

[91] Alexandros Nikou, Jana Tumova, and Dimos V. Dimarogonas. 2017. Probabilistic plan synthesis for coupled multi-agent systems. *IFAC World Congr.* 50, 1 (2017), 10766–10771. DOI: https://doi.org/10.1016/j.ifacol.2017.08.2280

[92] Reza Olfati-Saber and S. Shamma, Jeff. Consensus filters for sensor networks and distributed sensor fusion. In *Proceedings of the 44th IEEE Conference on Decision and Control.* 6698–6703. DOI: 10.1109/CDC.2005.1583238

[93] Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 10 (2010), 1345–1359. DOI: https://doi.org/10.1109/TKDE.2009.191

[94] Sue T. Parker, Robert W. Mitchell, and Maria L. Boccia. 2006. *Self-awareness in Animals and Humans: Developmental Perspectives.* Cambridge University Press.

[95] Judea Pearl and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect.* Basic Books.

[96] Jean Piaget. 2005. *Language and Thought of the Child: Selected Works*, Vol. 5. Routledge.

[97] Claudio Piciarelli, Lukas Esterle, Asif Khan, Bernhard Rinner, and Gian L. Foresti. 2016. Dynamic reconfiguration in camera networks: A short survey. *IEEE Trans. Circ. Syst. Vid. Technol.* 26, 5 (May 2016), 965–977. DOI: https://doi.org/10.1109/TCSVT.2015.2426575

[98] Steven Pinker. 1999. How the mind works. *Ann. N. Y. Acad. Sci.* 882, 1 (1999), 119–127.

[99] Jeremy Pitt. 2014. *The Computer after Me: Awareness and Self-awareness in Autonomic Systems.* World Scientific, Singapore.

[100] Marco Platzner. 2016. On-the-fly computing: Self-aware heterogeneous multi-cores. In *Proceedings of the International Conference on Hardware/Software Codesign and System Synthesis.* 1–2.

[101] Athanasios S. Polydoros and Lazaros Nalpantidis. 2017. Survey of model-based reinforcement learning: Applications on robotics. *J. Intell. Robot. Syst.* 86, 2 (01 May 2017), 153–173. DOI: https://doi.org/10.1007/s10846-017-0468-y

[102] Jürgo S. Preden, Kalle Tammemäe, Axel Jantsch, Meiro Leier, Andri Riid, and Emine Calis. 2015. The benefits of self-awareness and attention in fog and mist computing. *Computer* 48, 7 (2015), 37–45. DOI: 10.1109/MC.2015.207

[103] Steven R. Quartz and Terrence J. Sejnowski. 1997. The neural basis of cognitive development: A constructivist manifesto. *Behav. Brain Sci.* 20, 4 (1997), 537–556.

[104] D. Ratasich, F. Khalid, F. Geissler, R. Grosu, M. Shafique, and E. Bartocci. 2019. A roadmap toward the resilient internet of things for cyber-physical systems. *IEEE Access* 7 (2019), 13260–13283. DOI: https://doi.org/10.1109/ACCESS.2019.2891969

[105] Mahdyar Ravanbakhsh, Mohamad Baydoun, Damian Campo, Pablo Marin, David Martin, Lucio Marcenaro, and Carlo S. Regazzoni. 2018. Learning multi-modal self-awareness models for autonomous vehicles from human driving. In *Proceedings of the 21st International Conference on Information Fusion.* 1866–1873. DOI: https://doi.org/10.23919/ICIF.2018.8455667

[106] Mahdyar Ravanbakhsh, Mohamad Baydoun, Damian Campo, Pablo Marin, David Martin, Lucio Marcenaro, and Carlo S. Regazzoni. 2018. Hierarchy of gans for learning embodied self-awareness model. In *Proceedings of the 25th IEEE International Conference on Image Processing.* 1987–1991. DOI: https://doi.org/10.1109/ICIP.2018.8451418

[107] Brien Riley. 2006. The birth of the mind: How a tiny number of genes creates the complexities of human thought. *J. Nerv. Ment. Dis.* 194, 3 (2006), 229.

[108] Philippe Rochat. 2003. Five levels of self-awareness as they unfold early in life. *Conscious. Cogn.* 12, 4 (2003), 717–731.

[109] Andreas Roepstorff, Jörg Niewöhner, and Stefan Beck. 2010. Enculturing brains through patterned practices. *Neural Netw.* 23, 8–9 (2010), 1051–1059.

[110] Matt Rosencrantz, Geoffrey Gordon, and Sebastian Thrun. 2003. Decentralized sensor fusion with distributed particle filters. In *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence.* 493–500.

[111] Stuart J. Russell and Peter Norvig. 2016. *Artificial Intelligence: A Modern Approach.* Pearson Education Limited.

[112] Paul Ruvolo and Eric Eaton. 2013. Active task selection for lifelong machine learning. In *Proceedings of the AAAI Conference on Artificial Intelligence.*

[113] Santanu Sarma, Nikil Dutt, Puneet Gupta, Nalini Venkatasubramanian, and Alexandru. Nicolau. 2015. CyberPhysical-system-on-chip (CPSoC): A self-aware MPSoC paradigm with cross-layer virtual sensing and actuation. In *Proceedings of the Conference Exhibition on Design, Automation Test Europe.* 625–628. DOI: https://doi.org/10.7873/DATE.2015.0349

[114] Julia Schaumeier, Jeremy Pitt, and Giacomo Cabri. 2012. A tripartite analytic framework for characterising awareness and self-awareness in autonomic systems research. In *Proceedings of the IEEE 6th International Conference on Self-Adaptive and Self-Organizing Systems Workshops.* 157–162. DOI: https://doi.org/10.1109/SASOW.2012.35

[115] Fedor Shmarov, Sadegh Soudjani, Nicola Paoletti, Ezio Bartocci, Shan Lin, Scott A. Smolka, and Paolo Zuliani. 2019. Automated synthesis of safe digital controllers for sampled-data stochastic nonlinear systems. *CoRR* abs/1901.03315 (2019).

[116] Daniel L. Silver, Qiang Yang, and Lianghao Li. 2013. Lifelong machine learning systems: Beyond learning algorithms. In *Proceedings of the AAAI Spring Symposium Series*.

[117] Khim L. Sim and John W. Rogers. 2008. Implementing lean production systems: Barriers to change. *Manage. Res. News* 32, 1 (2008), 37–49.

[118] Adam Smith. 2006. Cognitive empathy and emotional empathy in human behavior and evolution. *Psychol. Rec.* 56, 1 (2006), 3–21.

[119] Demetri P. Spanos, Reza Olfati-Saber, and Richard M. Murray. 2005. Distributed sensor fusion using dynamic consensus. In *Proceedings of the International Federation of Automatic Control World Congress (IFAC'05)*.

[120] Scott D. Stoller, Ezio Bartocci, Justin Seyster, Radu Grosu, Klaus Havelund, Scott A Smolka, and Erez Zadok. 2012. Runtime verification with state estimation. In *Runtime Verification*, Sarfraz Khurshid and Koushik Sen (Eds.). Springer, Berlin, 193–207.

[121] Richard S. Sutton, Andrew G. Barto, et al. 1998. *Introduction to Reinforcement Learning*. Vol. 135. MIT Press, Cambridge, MA.

[122] Likewin Thomas, Kiran Manjappa, B. Annappa, and G. Ram Mohana Reddy. 2011. Parallelized K-means clustering algorithm for self aware mobile ad-hoc networks. In *Proceedings of the 2011 International Conference on Communication, Computing and Security*. ACM, 152–155. DOI: https://doi.org/10.1145/1947940.1947973

[123] Sebastian Thrun and Lorien Pratt. 2012. *Learning to Learn*. Springer Science & Business Media.

[124] Stéphane Tirard. 2015. *Abiogenesis*. Springer, Berlin, 25–25. DOI: https://doi.org/10.1007/978-3-662-44185-5_2

[125] Lisa Torrey and Jude Shavlik. 2010. *Transfer Learning*. IGI Global, 242–264.

[126] Phil Turner. 2005. Affordance as context. *Interact. Comput.* 17, 6 (07 2005), 787–800.

[127] Edward Burnett Tylor. 1871. *Primitive Culture: Researches into the Development of Mythology, Philosophy, Religion, Art, and Custom*. Vol. 2. J. Murray.

[128] Sebastian Uchitel, Greg Brunet, and Marsha Chechik. 2007. Behaviour model synthesis from properties and scenarios. In *Proceedings of the 29th International Conference on Software Engineering*. IEEE Computer Society, 34–43.

[129] Johan Van Benthem. 2011. *Logical Dynamics of Information and Interaction*. Cambridge University Press, Cambridge, UK.

[130] Hans Van Ditmarsch, Wiebe van Der Hoek, and Barteld Kooi. 2007. *Dynamic Epistemic Logic*. Vol. 337. Springer Science & Business Media.

[131] Anh Vu and Jay A. Farrell. 2015. Feature mapping and state estimation for highly automated vehicles. *J. Contr. Dec.* 2, 1 (2015), 1–25. DOI: https://doi.org/10.1080/23307706.2015.1007565

[132] Lev Semenovich Vygotskiĭ. 2012. *Thought and Language*. MIT Press, Cambridge, MA.

[133] Sandra Wachter, Brent Mittelstadt, and Luciano Floridi. 2017. Transparent, explainable, and accountable AI for robotics. *Sci. Robot.* 2, 6 (2017). DOI: https://doi.org/10.1126/scirobotics.aan6080

[134] Eric K. Wang, Yunming Ye, Xiaofei Xu, Siu-Ming Yiu, Lucas Chi Kwong Hui, and Kam-Pui Chow. 2010. Security issues and challenges for cyber physical system. In *Proceedings of the IEEE/ACM International Conference on Green Computing and Communications International Conference on Cyber, Physical and Social Computing*. 733–738. DOI: https://doi.org/10.1109/GreenCom-CPSCom.2010.36

[135] Christopher Wing. 2016. Better keep your hands on the wheel in that autonomous car: Examining society's need to navigate the cybersecurity roadblocks for intelligent vehicles. *Hofstra L. Rev.* 45, 2 (2016), 707–vi.

[136] Wayne Wolf. 2009. Cyber-physical systems. *Computer* 42 (2009), 88–89. DOI: 10.1109/MC.2009.81

Received November 2018; revised July 2019; accepted November 2019