Advanced Robotics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tadr20

What is the role of the next generation of cognitive robotics?

Shingo Shimoda, Lorenzo Jamone, Dimitri Ognibene, Takayuki Nagai, Alessandra Sciutti, Alvaro Costa-Garcia, Yohei Oseki & Tadahiro Taniguchi

To cite this article: Shingo Shimoda, Lorenzo Jamone, Dimitri Ognibene, Takayuki Nagai, Alessandra Sciutti, Alvaro Costa-Garcia, Yohei Oseki & Tadahiro Taniguchi (2022) What is the role of the next generation of cognitive robotics?, Advanced Robotics, 36:1-2, 3-16, DOI: 10.1080/01691864.2021.2011780

To link to this article: https://doi.org/10.1080/01691864.2021.2011780

ADVANCED ROBOTICS
2022, VOL. 36, NOS. 1–2, 3–16
https://doi.org/10.1080/01691864.2021.2011780

SURVEY PAPER

OPEN ACCESS Check for updates

What is the role of the next generation of cognitive robotics?

Shingo Shimoda[a], Lorenzo Jamone[b], Dimitri Ognibene[c], Takayuki Nagai[d], Alessandra Sciutti[e], Alvaro Costa-Garcia[a], Yohei Oseki[f] and Tadahiro Taniguchi[g]

[a]RIKEN Center for Brain Science TOYOTA Collaboration Center, Nagoya, Japan; [b]School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK; [c]Computer Science and Artificial Intelligence, University of Milano Biccoca, Milano, Italy; [d]Graduate School of Engineering Science, Osaka University, Osaka, Japan; [e]Cognitive Architecture for Collaborative Technologies Unit, Italian Institute of Technology, Genova, Italy; [f]Department of Language and Information Sciences, University of Tokyo, Tokyo, Japan; [g]Department of Human and Computer Intelligence, Ritsumeikan University, Shiga, Japan

ABSTRACT

Social demand for robots to be our partners in daily life has been rapidly increasing. Cognitive robotics should play a major role in making robots our partners. To discuss the role of cognitive robotics, we organized the round table in December 2020. This review paper aimed at clarifying the role of cognitive robotics summarizing the discussion in the round table. The round table noted that the existence of uncertainty in the continuous control loop is a source of the need for cognitive robots and is the key factor that distinguishes cognitive robotics from the cognitive system in other fields. This paper summarized the discussion focusing on the creation of several cognitive functions without stopping even if the robots face novel uncertainty in daily life. We discussed information generalization, active sensing, prediction, and language communication as the necessary functions for future cognitive robots. One of the conclusions of the discussion is the importance of setting primitive but concrete targets for cognitive robotics research as cognitive robotics problems. We should continue to discuss the setting of these targets as a grand challenge for cognitive robotics.

ARTICLE HISTORY
Received 31 May 2021
Revised 19 September 2021
Accepted 16 October 2021

KEYWORDS
Cognitive robot; information generalization; active sensing; prediction; language communication

1. Introduction

What is the role of cognitive robotics? Today, the line between automatic machines and what people commonly call robots is becoming increasingly blurred. However, robots were originally conceived as artifacts that move in a more biological or animal-like manner. Therefore, many robotics researchers are seeking to develop robots that can live with us and help each other.

Recently, the social demand for robots to become our partners has been rapidly increasing. For this purpose, robots need to perceive the environment, understand the surrounding circumstances, communicate with people, and move safely sharing the same environment with humans. Although the abstract argument that 'a high level of cognitive function is essential for those robots' is not in dispute, the kinds of required cognitive functions remain an open question.

In psychology, cognition refers to the process of acquiring knowledge and understanding through the senses, thought, and experiences [1,2], implying that cognition is mainly a one-way process from the perception to the understanding of the environment. In the fields of behavioral biology [3] and artificial intelligence [4], which focus more on dynamic problems, the meaning of cognition has become broadened to include not only the one-way process of understanding a given situation, but also the process of creating motion with the functions of learning, memory, and motion controls as a signal loop. We are very curious about how robotics will contribute to improving the concept of cognition.

In 2014, the Technical Committee of Cognitive Robotics (TC-CoRo) [5] was established at the IEEE Robotics and Automation Society to encourage discussion of cognitive robotics research. Prof. Giulio Sandini, founding principal chair of TC-CoRo, stated at the time when TC-CoRo was launched that cooperation among various fields of science is essential for realizing a truly useful cognitive system, and that robotics need to serve as a 'melting pot' for integrating these sciences.

This message arises from the important feature of robotics that all of the functions implemented in a controller are embodied as robot behaviors in the real world. We can observe and analyze the details of the implemented functions in the behaviors beyond the theoretical and abstract discussions. This feature of robotics prevents the discussions of the conceptual functions from

CONTACT Shingo Shimoda shingo.shimoda@riken.jp

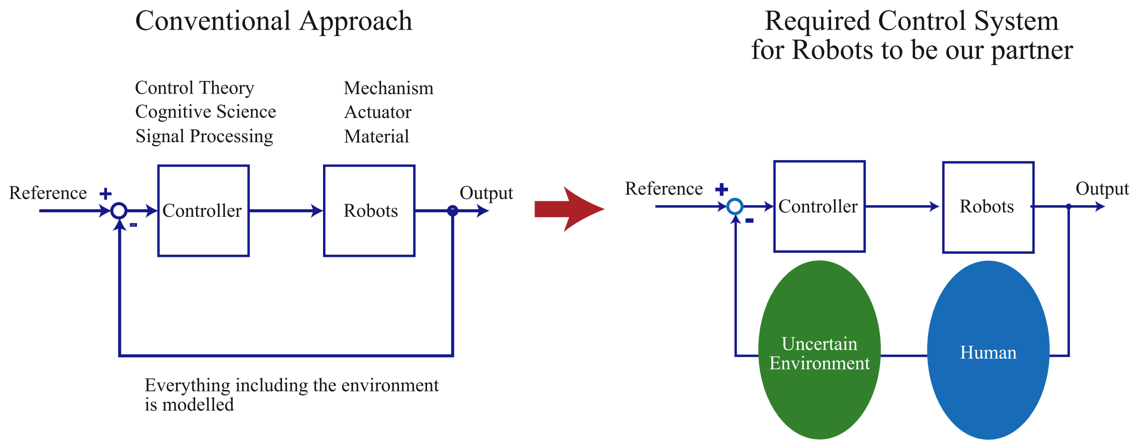

Figure 1. Conventional approach: Everything is modeled and the controller is designed to fit the model. Uncertainties such as modeling error and measurement noise are dealt with by the robustness of the controller. Required control system: Higher level uncertainties such as behavior targets in terms of context and communication with humans exist in the control loop. In this case, cognitive systems are needed for robot control.

becoming too diverse and too general. Consider this feature in terms of the contribution of robotics to improve the concept of cognition, robotics can provide a field to connect the computation and the behaviors with physical contacts with environment and communication with others, where we can discuss in-depth cognitive functions including the existence of cognition in the process and the role of cognition.

In order to clarify this role of robotics, we held an on-line round table entitled 'What is the role of the next generation of cognitive robotics?'. In the round table, we discussed the topic ranging from the basic skills for cognitive functions to implementation in the real world (all discussions are available at [6]).

In this paper, we summarize the discussions from the round table to clarify the appropriate cognitive system for robot control and the proper direction of cognitive robotics research in the future. To summarize the discussions from the round table, we first need to clarify why cognitive systems are necessary from a robotics point of view. In classical robotics, robots are controlled using a system as shown in Figure 1(a). In this framework, important information for control such as the robot bodies, the environment, behavioral goals and constraint conditions are modeled, and a controller matching to the model is designed. One of the critical problems of this model-based approach is the error from the model. Even in a stable environment such as in a factory, as long as the robot moves and that movement is measured by sensors, uncertain modeling errors and sensor noise are inevitable. The common method for dealing with these uncertainties is to design robustness into the controller. Through the many discussions for modeling methods and robust controller design, the model-based approach has been a great success for creating robots that have abilities superior to humans in quick and accurate repetitive tasks in stable environments, as exemplified by industrial robots.

However, as the applications of robots have become more diverse, it has become increasingly difficult to deal with the uncertainties by only controller robustness. Especially in daily life, robots face uncertainties that are qualitatively different from those in a factory, such as situations in which the environment is always changing or behavioral goals can only be set in context. Interactions with humans including communication using ambiguous expressions are also important uncertainty in daily life. These types of uncertainties cannot be dealt with by the conventional robustness of the controller. At this stage, we clearly recognize that cognitive functions are needed for robots beyond the conventional robustness.

In the round table discussion, in addition to the high class of uncertainties, we considered continuous control as another important feature of the cognitive system in robotics. In order for a robot to be a partner in daily life, it must manage uncertainties in appropriate ways without stopping control, even if the uncertainty is a completely novel one.

These two points described in Figure 1(b) can be the key problems of cognitive robotics. In other words, the capability of overcoming novel uncertainties online through robot body–environment interactions distinguishes cognitive robotics from cognitive systems in other fields.

What are the key issues for continuous control with uncertainties? In the round table, several important candidates for the key issues were raised, namely, generalization, active sensing, prediction and language communication. In this review paper, we discuss these topics focusing on continuous control with uncertainties. As

Figure 2. The objects and robots used in the experiments of [8,9]. The image on the left shows 500 objects. The robot is shown at the top right. The bottom right shows the human-labeled categories of the objects used in the experiments.

mentioned above, the approaches to establish these functions without stopping robot control in the real world is the key problem for cognitive robotics. The discussions in this paper review the various system for cognitive robotics from this point of view and derive a future direction for research on cognitive robotics. Artificial intelligence based on machine learning is also an important tool for cognitive robotics. How to use various types of tools in continuous control is also an important target of discussion.

In Section 2, we first discuss the importance of and the problems associated with environment information generalization and derive an important approach for using generalized information. We also discuss the role of active sensing for information generalization. In Section 3, we show the importance of prediction for robots beyond the conventional dynamics-based prediction when living with humans in the same environment. In Section 4, we deepen our discussion on the possibilities that language communication between humans and robots creates. In Section 5, we conclude this paper by showing the importance of building common targets for cognitive robotics problems.

2. Environment understandings

2.1. Importance of information generalization for cognitive robotics

How should a robot handle the environmental information when the environment becomes complex? The concept of 'the complexity of the environment creates the complexity of the behavior ', as represented by Subsumption Architecture [7] in which various reactive behaviors are designed *a priori*, has many implications. However, as another important approach, many cognitive robotics researchers will agree with the idea that generalization of the environmental information is an important solution for robots to understand the environment and move in complex environments. Moreover, when the complexity of the environment includes ambiguity in behavioral goals and communication with humans, generalization of the environmental information is necessary for robots. In this section, therefore, we will summarize the several approaches of information generalization to understand the environment and discuss the appropriate functions for robot controls in the complex environment.

2.2. Labeling for environment understanding

Labeling of the surrounding environment is an important first step toward understanding the environment. Several approaches have been discussed for labeling the environment with robots. Nishihara et al. have succeeded in grounding about 100 words to more than 500 objects shown in Figure 2 through robot-human interaction [8,9]. This study is based on the idea that object categories can be learned unsupervised by co-occurrence of object multimodal features. In other words, by clustering multimodal features, it is possible to extract the main features that can represent the category and discard the features that are not so relevant.

This result can be said to realize generalization. In addition, linguistic information given by humans is also an important clue for forming categories. However, since robots do not have linguistic knowledge such as a lexicon in advance, that linguistic knowledge cannot be used immediately. Robots need to segment human utterances and learn what phonological patterns exist to acquire words. The authors solved the problem of finding the connection between them while learning this word and the multimodal features obtained from the object at the

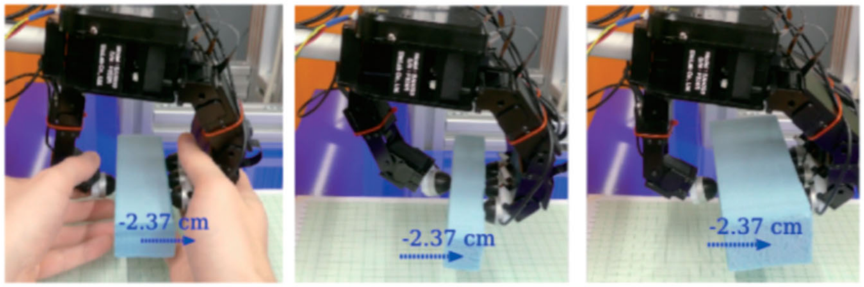

Figure 3. A human user is teaching an in-hand manipulation action on a specific object. The robot can then generalize the learned action on different objects, e.g. smaller or larger objects. Adapted from [11].

same time. This mimics the process by which a baby, who cannot understand the meaning of words at all, gradually learns the names of objects statistically through interactions with his parents. Furthermore, Miyazawa et al. extended this idea to robot action learning [10]. This showed that robots can generalize objects, their names, and actions and connect them to each other at the same time.

2.3. Sensitivity and insensitivity to the environment for cognitive robotics

In the process of generalization of environmental information, computation need to be 'insensitive' to the small environmental changes. For instance, the teddy bear in Figure 2 needs to be labeled as a 'teddy bear' even if a small change occurs such as becoming slightly broken or dirty. The capability for insensitivity to small changes is essential for a stable understanding of the environment.

However, understanding the environment in a stable way is not always an advantage. Stable understanding may reduce the diversity of robot behaviors in response to the complexity of the environment. This fact implies that the robot behavior will become uniform and lack the ability to adapt to environmental changes if the robot uses only generalized information. The controller thus requires aspects of both high and low sensitivity to the environment.

Several approaches have been proposed for balancing stable understanding and reactive behavior control. Solak et al. proposed a robust compliant controller for dexterous in-hand manipulation by combining Dynamical Movement Primitives for generalized movement with Virtual Springs Framework for real-time feedback of the contact forces measured on the robot fingertips [11]. They experimentally succeeded with in-hand translation and rotation of unknown objects as described in Figure 3. Choi et al. discussed the object handling problem by combining controllers for reactive motion and for the target posture setting [12,13]. This approach considered realistic sensor modalities with reasonable delays in daily life for each controller to create appropriate behaviors, thereby creating dexterous handling motions.

Recently, deep reinforcement learning has been used to learn in-hand manipulations of a Rubik's cube with a dexterous robotic hand [14]. Although it is impressive to see how interesting strategies (e.g. finger gaiting, multi-finger coordination, controlled use of gravity) would naturally emerge in the robot behaviors to solve the task, the approach requires to collect a large amount of annotated data and the robot was not required to hold the object against gravity with the fingers while manipulating it (i.e. the Rubik's cube is held on top of the palm of the robot hand). Interestingly, in [15] the reinforcement learning procedure is combined with a low-level reactive controller based on tactile feedback, permitting to learn more complex in-hand manipulation tasks while minimizing the amount of failures during learning. However, none of these works have shown generalization to different objects being manipulated. The work in [11] achieves generalization to new objects by representing the object movement using a virtual reference frame attached to the fingertips, based on the Virtual Springs Framework. By using this representation, in-hand manipulation actions are learned from human demonstrations using Dynamical Movement Primitives and then executed with a compliant reactive controller that uses feedback of the contact forces measured on the robot fingertips. The reported experiments show that learned in-hand motions (e.g. arbitrary translations and rotations) could be then executed by the robot on different objects, as described in Figure 3.

The combination of generalized information and reactive control is useful not only in in-hand manipulation but also in mobility control. Okajima et al. proposed [16] bipedal walking control using generalized behavior goals with reactive behavior tuning by tacit learning, which is a behavior-based adaptation architecture [17]. They successfully turned the walking direction by changing the simple signals that represent the motion intentions. Control loops of different frequencies in a single control system can contribute to balancing the stability and reactivity to environmental information. Miyazaki et al. showed the stability of the bipedal walking control dividing it into the two modes for singular perturbation analysis [18]. They discussed the motion of the center of gravity in the slow mode subsystem as a global locomotion factor, while each joint motion was dealt with in a fast mode subsystem. Taniguchi and Nagai et al. demonstrated that mobile robots can move around for the purpose of shopping while symbolizing the objects in the shop [19].

These research results suggest that one of the important advantages of stable understanding by generalizing environmental information is the autonomous setting of behavior goals in the continuous control loop while using reactive motions to adopt the behaviors to complex environmental changes. The simultaneous use of both functions with sensitive and insensitive to environmental changes is likely to become more mainstream for adapting to the environmental changes and for understanding the environment by using generalized information.

2.4. Active sensing for effective perception

To achieve the generalization of information, the choice of information, that is, which information to use and which to discard, is also an important issue. When we consider human senses, we are not only insensitivity to environmental changes but we also unconsciously ignore some of the environmental information for the purpose of generalization. In the round table discussion, we posed the question 'When to stop sensing and start acting ?' and discussed what combination of sensing and action initiation is needed to produce appropriate sensing in a complex environment. In this context, Prof. Dimitri Ognibene mentioned the importance of active sensing. He pointed out the limitations of the immediate and full perception of any non-trivial environment because of the sheer amount of data flowing in the sensors from the environment even though we succeeded in creating well-organized information generalization system. At the same time, sensory limitations, such as occlusions, limited resolution, signal-to-noise ratio, and others impede achieving the perception of the environment even with state-of-the-art sensors and algorithms [20,21].

By extending control to robots' own sensors, active vision [21–26] represents a biologically inspired strategy for dealing with sensory limitations. Directing the sensors allows the robot to become sensitive to the relevant parts of the environment, which may have been inaccessible till that moment, while obtaining insensitivity to the irrelevant ones that become inaccessible after the change of sensory configuration, e.g. get out of the field of view.

Active vision can then be seen as a physical embodiment of the interaction between sensitivity and insensitivity discussed before. By transferring parts of the computational implementation of this complex interaction to the physical level, active vision can have such advantages that evolution appears to have etched it into the anatomy of the human eye. Indeed, the eye has a high-resolution area, fovea, that can maximize sensitivity to relevant information and a low-resolution periphery that helps directing the fovea [25,27]. Humans' eye control also reflects the contextual optimization of sensitivity and insensitivity: in fact humans follow different sensing strategies when performing different tasks and access 'on demand' the information most relevant for the current state of the task [28,29].

However, this computational simplification is just a supplementary advantage of active perception systems that tackle the imperative of minimizing uncertainty under unavoidable sensory limits [30]. While in some relatively simple environments useful active perception strategies can be implemented even by reactive controllers, i.e. direct sensor-action mappings [31,32], in the general case, active perception [33] may require representations and computations [34] more complex than those usually found in typical robot control tasks. This is caused by the necessity to accumulate useful information that could become inaccessible at the next time step [27,35–38], and predict the next most informative actions [39,40], all of which can be computationally hard [41,42]. Considering its computational and representational demands as well as its crucial role for the employment of robots in non-trivial contexts [37,43,44] active perception may be one of the main reasons to develop 'cognitive' robots [45].

'Deciding when to stop sensing and start acting' or, in other words, the trade-off between gathering more information or greedily performing the current best action, which can also be seen as a formulation of the famous exploration-exploitation dilemma [46,47], is an important aspect of active perception. This aspect becomes even more important when considering social interaction and learning, two central requirements for cognitive robots' applications.

Another challenge comes from learning under active perception conditions. Recently, several methods to learn active perception skills in simulated environments with realistic stimuli have been presented [37,43,44,48]. However, these approaches often permit several types of simplifications leaving how to learn new active perception skills in real environments still an open question. Indeed, learning to overcome sensory limitations has been shown to present several theoretical issues [49]. Yet some experimental results show that both attention and active perception may largely speed up learning and improve generalization by exploiting their ability to be insensitive to irrelevant stimuli [25,50].

2.5. Future direction of information processing for cognitive robotics

As Prof. Doyle pointed out, biological control systems can be represented by Bow-tie structure [51], in which the dimensionality of the environmental input signals is first reduced before several signal processing is performed using the low-dimensional signals [52,53], and then the signals are again taken back to a higher dimensionality to achieve adaptive motor control in a complex environment [54–57]. In biological control systems, it seems that different levels of generalized signals are used for different purposes from higher brain functions such as behavior goal setting, decision making to the low-level behavior controls such as force control, motion skill tuning [58] and sensor integration. A biological system can overcome novel uncertainties in continuous control using these functions. Although the biological approach is not the only solution, it is definitely a method that we should learn from. We need to accelerate the discussion on how to observe and analyze signals for more adaptive and stable behavioral control with understanding the complex environment.

3. Behavior prediction

3.1. Appropriate prediction for cognitive robotics

One of the most important benefits of robots to be able to use generalized information well may be the prediction of future events. Prof. Sandini, founding principal chair of TC-CoRo, proposed the concept of 'Beyond Real-time' [59] for cognitive robotics, suggesting that the ability to move with predictions of future events beyond real-time responses is essential for future cognitive robotics. He stated that this capability is expected to lead to the notion of cognitive safety, where humans and robots can live in the same space with a high level of safety if the robots can predict what the human will do or want to do [60].

How are robots able to make prediction in the continuous control loop while understanding the environment? What is the critical difference from the predictions in data science? The most important feature of the prediction for robots is that the environment and scene are treated as dynamics. This is strongly related to the concept of Embodied Cognition [61–63], where the robot body, movement, and interactions with the environment play a significant role in cognition. However, if robots are to become more active as our partners, 'dynamics' need to have a more human-like meaning rather than simply saying that they have bodies or movements.

3.2. Important dynamics for prediction of cognitive robotics

One of the important dynamics of the external environment for human-like cognition is changes in the meaning of objects. The notion of affordance, which is the concept that the environment affords us the meaning of objects, is the key concept for the dynamics of the environment in this scope. In terms of the importance of affordance in robot control, the review paper by Prof. Lorenzo Jamone [64] has inspired several works in robotics. Therefore, even though we do not examine it in depth here, we expect that the robots may make predictions such as, 'He will sit down on this table' when affordance is well implemented in the robot controller.

Another important type of dynamics for cognition may be the internal state of the robot according to the situation as well as the interpretation of the external environment. At the round table, Dr Alessandra Sciutti raised the issue on the importance of 'embodied communication'. There are a variety of movements that humans perform and process unconsciously in other agents. For instance, when people look around, their eyes immediately reveal where the focus of their visual attention is, and it is easy to predict which object they will most probably take from gaze analysis. Other examples of human ability to understand implicit signals are the possibility to recognize someone's emotional state from subtle changes in their facial expression [65,66], their voice [67], or even their body movements [68] or infer their attention or arousal from a change in their pupil size [69]. These kinds of predictions are made possible by the motor regularities which all humans share in their motor repertoire [70]. Implicit signals are so important for interaction because they represent the backbone of 'emergent coordination' [71]: mutual adaptation, synchronization, and anticipation which occur without awareness and drastically reduce the cognitive load and delays in interactions.

The importance of these embodied communications has been recognized recently not only in cognitive science and neuroscience but also in robotics and AI [72]. Robot cognition needs to allow for understanding the movements of the human partners to enable anticipation of their intentions. Moreover, robots needs to be able to plan motions transmitting similar signals [73] to make the robot actions intuitively interpretable and

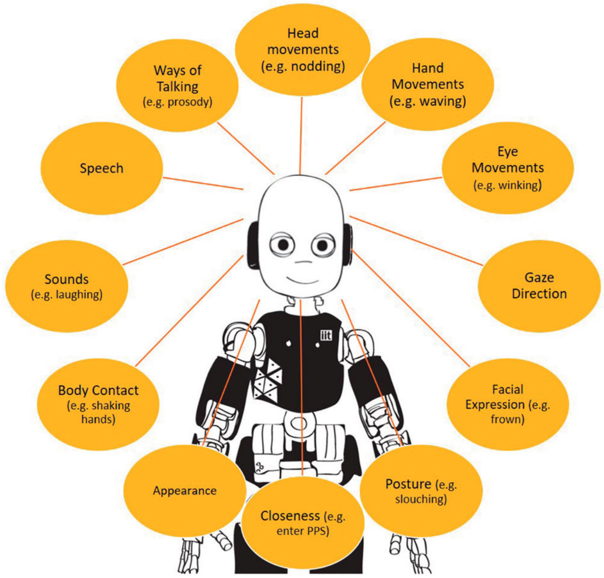

Figure 4. Human-robot mutual understanding needs to leverage the exchange and comprehension of a variety of embodied signals. They are usually processed subconsciously by the brain during human-human interaction, enabling fast and effective collaboration.

legible to facilitate effective collaboration with humans (Figure 4) [74].

These behaviors are known to be strongly related not only to emotions but also to the brain higher functions such as Sense of Agency (SoA) [75] and Time Perception (TP) [76,77], which have been attracting attention in recent years for understanding humans' behavior, and these internal states are also concepts that can be discussed only in a continuous control system.

Although it is not known exactly what role plays in humans, it is certain that they underpin the higher brain functions of cognition and decision that lead to consciousness. Is it necessary for the cognitive system of a robot to have such functions? Although the importance of SoA in robot behavior has not been fully discussed, the merit of quantifying SoA for robots can be developed into a discussion of human-robot collaborations. For instance, Ueda et al. [78] studied the relationship between the amount of assistance and SoA of subjects during an operation to quantify the appropriate assistance level for autonomous driving systems. They showed that a phenomenon similar to the so-called 'uncanny valley' also occurs in the perception of SoA. In their experiments, SoA increased up to a certain point as the amount of assistance increased because the target object moved more in accordance with intentions of the subject. However, when the amount of assistance exceeded a certain level, the subject began to feel a sense of discomfort losing SoA.

As described in the paper [78], it is clear that humans do not feel comfortable with excessive support by robots though the appropriate level of robot support for humans has not yet been clarified. Recently, mathematical models of SoA have been proposed [79], and we can say that the groundwork for discussing SoA for robots has been laid. To help each other sharing appropriate SoA between humans and a robot could be one of the ideal visions for the future.

3.3. Prediction beyond reality

Dr Sciutti introduced the appealing phrase 'Beyond Reality' in the round table, suggesting that robots should not perceive the world as it is but should have perceptual and cognitive biases for collaboration with humans. It is well known that human brains transform 'real' sensory signals into the information what we want to sense, which is sometimes called 'anthropomorphic lens'. As a result, perception of the temporal and spatial properties of actions or the environment can be inaccurate [80], as demonstrated for instance by visual illusions. However, these processes bring several advantages for collaboration. For instance, when observing a passing action, the human doesn't detect the details of the motion, but naturally understands the features of the action such as the action goal [81], which is more important than the details for good collaboration. The notion of beyond reality, therefore, could be an important index for information generalization discussed in the previous section.

4. Language communication

4.1. Language for cognitive robotics

Language communication is one of the most important higher-order cognitive functions, as also highlighted in the round table. Our social activities would not be possible without language communication, which allows us to use abstract expressions without detailed definitions. The ability is learned through communication with people, beginning with reflexive voice responses and two-words sentences in infants, and then naturally developing the grammatical rules. Surprisingly, depending on how and with whom we communicate in infancy, we can make any language our mother tongue. The robots, therefore, must learn and use languages to integrate into human society.

Prof. Oseki pointed out that, in order to build cognitive robots that process and learn natural languages like people [82], we should 'reverse-engineer' human language processing and learning, as advocated in the computational cognitive science literature. Specifically, in the computational cognitive science of language, computational models of language processing and learning are constructed from symbolic generative models and artificial neural networks originally developed in natural language processing (NLP), and then evaluated against human behavioral and neuroimaging data experimentally measured in cognitive and brain sciences. The key idea here is that the fusion between the science-oriented 'symbolist' approach and the engineering-oriented 'connectionist' approach to language processing and learning will be important for the next generation of cognitive robotics.

Prof. Oseki also presented recent results that demonstrate that bigger state-of-the-art models called Transformers [83], despite impressive performances via engineering evaluation metrics, are not always 'human-like' relative to smaller counterparts trained on less training data [84], and symbolic-neural architectures called Recurrent Neural Network Grammars (RNNGs) [85] outperform Long Short-Term Memory (LSTM) baselines and, most importantly, generalize better to unseen linguistic environments [86]. For future research, given that language processing is an instance of information processing ultimately realized in the human brain, the 'hardware implementation' level in Marr's three levels of description [87] must be integrated with the computational cognitive science of language processing and learning towards the computational cognitive neuroscience [88] in order for cognitive robots to communicate with humans in uncertain environments.

4.2. Cognition through language learning and understanding

Language learning and understanding are also important cognitive capabilities that robots should have. Prof. Taniguchi pointed out the important feature of the role of human language that distinguishes language communication of human beings from that of other animals. Human cognition is adaptive not only adaptive to the physical environment but also to the semiotic environment. In many conventional discussions in cognitive robotics, the terms 'symbols' and 'symbolization' have been used naively to refer to internal representation related to words. The discussion has been influenced by the physical symbol systems hypothesis where symbols are characterized as just a discrete token [90]. The notion of symbols in semiotics, e.g. semiosis in Peirce's semiotics, was ignored. Symbol systems should be regarded as a dynamic system involving emergent property, i.e. a symbol emergence system [91]. A robot needs to have a cognitive capability to adapt to a symbol emergence system to become able to communicate with people in a long-term manner. This means a robot needs to have language learning and understanding capability at least [92]. Therefore, Prof. Taniguchi emphasized the importance of adaptability in semiotic communication in cognitive robotics.

Language is based on multimodal information, and cognitive systems integrating sensor-motor information will be a key for language communication. Learning a language is not just a problem of dealing with text data. We human beings understand the meanings of language in relation to the real-world sensor-motor information and generalized concept we form based on embodied physical experiences and semiotic communications. For example, when we try to understand a sentence, 'please go to the kitchen and bring me a bottle of water' and conduct the requested behavior, we need to relate the words to a specific object, place, and behavior. This means the wide range of language understanding, actually, is based on real-world multimodal sensor-motor information.

Therefore, developing integrative cognitive systems is a crucial step to realizing language communication in cognitive robotics. Prof. Taniguchi argued that a probabilistic generative model-based approach is promising. Taniguchi et al. proposed a multimodal spatial concept formation method called SpCoSLAM that integrates localization and mapping, image and speech recognition, spatial categorization, and lexical acquisition into a single probabilistic generative model (Figure 5) [93,94]. Recently, it becomes widely known that probabilistic modeling and inference is a general idea of machine learning. For example, the theory of control as probabilistic inference (CaI) shows that reinforcement learning can be regarded as an inference on a probabilistic generative model [95]. SpCoNavi is a navigation method based on SpCoSLAM and CaI [89]. SpCoNavi shows that learning spoken terms and understanding sentences in an indoor navigation task can be conducted purely based on robots' sensor-motor multimodal information.

4.3. Importance of language for further advances of cognitive robotics

For further progress, developing a totally adaptable integrative cognitive architecture is crucial not only in generalization, perception, and prediction but also in language communication. Prof. Taniguchi suggested that SERKET, a framework that can compose and decompose large-scale probabilistic generative models, is useful for developing future cognition of robots [19,96]. Developing a whole-brain probabilistic generative model is a future challenge [97] (Figure 6).

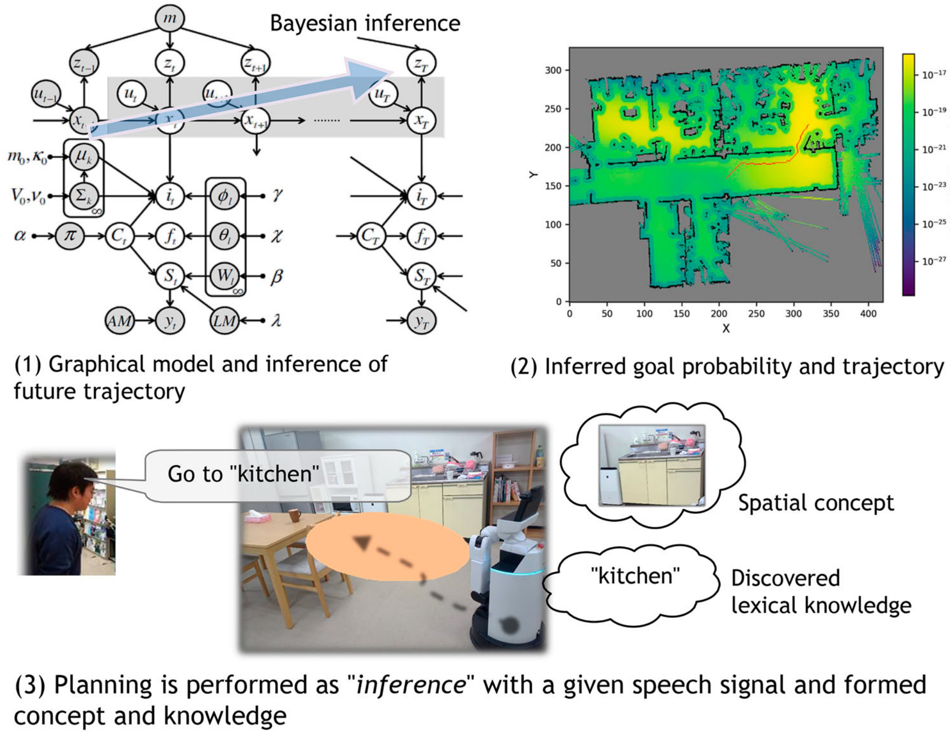

Figure 5. SpCoNavi can perform path planning using spatial concepts, vocabularies, a map learned in an unsupervised manner. The theory is consistent with model-based reinforcement learning based on the notion of CaI (control as probabilistic inference) [89].

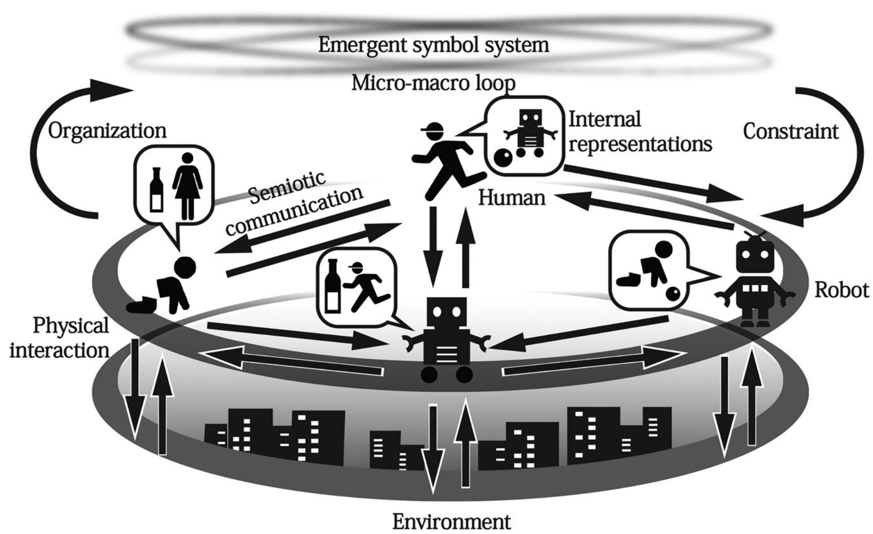

Figure 6. Overview of a symbol emergence system [92].

5. Conclusion for cognitive robotics problems

This paper reviewed the problems facing cognitive robotics based on discussions at a round table held in December 2020. The round table noted that the interdisciplinary discussions are necessary for further advances of cognitive systems and robotics can play the role of embodying the cognitive functions in a continuous control loop with uncertainty. In this review paper, we proceeded to discuss the necessary system for cognitive robots based on key functions of 'environment understandings with generalized information', 'active sensing', 'prediction of future event', and 'language communication'.

For environment understanding, we mentioned the importance of both high and low sensitivity to environmental inputs for achieving both stable understanding and adaptability to uncertainties. Information

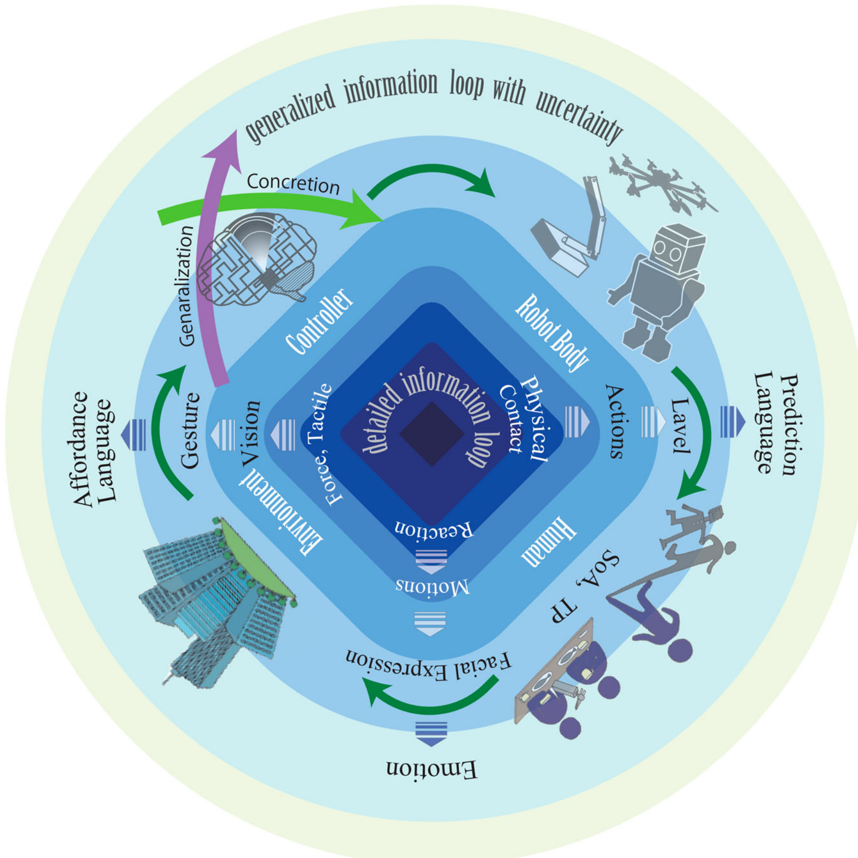

Figure 7. Control loops using generalized information can allow for temporal and semantic uncertainty and can handle language, prediction and affordances while control loops using embodied information focus on actual motion and dynamics.

generalization is an important approach for stable understanding, which will lead the specified functions such as autonomous behavior target setting.

In a complex environment, information generalization may not be enough to perceive all information. The robot should select the important and useful information by active sensing. In addition to the information selection, we clarified that active sensing will be useful for novel learning architecture beyond perception.

Prediction of future events is an important function that can be achieved based on environment understandings. Even though it is well known that prediction based on body dynamics is an important factor for robot control, more human-like prediction, which is sometimes progresses beyond reality, is necessary for a smooth collaboration with humans. These types of prediction emphasize the embodiment of robots compared with the prediction of data science that shows the event the most likely-to-happen.

Language communication is another key skill for future cognitive robots. Beyond information exchange, language between humans and robots needs to have adaptation with semiotic communication. We need further discussion on the computational model of language.

Though the physical interactions between the body and the environment always represent the real-time event, it is known that the information processing becomes more flexible to the uncertainties as the information is gradually generalized. This implies that we can keep remaining the uncertainties during continuous control when generalized information is used in the control loop. The uncertainties sometimes include time indeterminacy, suggesting that future event can be the target of control beyond reality. Therefore, the use of the various level of generalized information in a control loop is a necessary condition for controlling robots from the detailed motion control to normal life activities with us as described in Figure 7, and we believe that this type of control loop can be a melting pot of interdisciplinary discussions.

Through the discussions, the roles of the next generation of cognitive robotics are getting clear. We, human beings, can deal with uncertainties by sometimes responding reflexively and by sometimes behaving with uncertainties using generalized information. We do that by using various attractive functions such as affordance, SoA, and language communication. Cognitive robots must embody with their own approaches because robots have different bodies, different computational mechanisms, and different memory devices from humans, rather than aiming for those functions themselves.

In order to achieve the role of cognitive robotics, we think that it is important for setting primitive but concrete targets for cognitive robotics research. As mentioned in the introduction, discussion based on behavior in the real world is a significant feature of robotics. To encourage further advances in cognitive robotics, we need to set concrete behavior targets as needed for solving important cognitive robotics problems. We will continue to discuss the setting of these targets as a grand challenge for cognitive robotics.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Shingo Shimoda is a Unit Leader in Intelligent Behavior Control Lab. RIKEN Center for Brain Science TOYOTA collaboration Center. He received the BS, MS, and PhD in Space Robotics from the University of Tokyo, Japan, in 1999, 2001, and 2005, respectively. He spent as a visiting student with the Massachusetts Institute of Technology, USA, from 2003 to 2004. He was a Research Scientist in biomimetic intelligence at RIKEN from 2005 to 2007. In 2008, he became a Unit Leader in RIKEN Brain Research Institute in biomimetic intelligence and cognitive system. He is a principal chair of Technical Committee on Cognitive Robotics in IEEE Robotics and Automation Society and Steering committee member of IEEE Transactions on Cognitive and Developmental Systems. His research interest is a clarification of biological control system and their application for robot control and rehabilitation system.

Lorenzo Jamone is a Senior Lecturer in Robotics and Director of the CRISP group at the School of Electronic Engineering and Computer Science (EECS) of the Queen Mary University of London (QMUL). The CRISP group is part of ARQ (Advanced Robotics at Queen Mary). Since 2018 he is a Turing Fellow at The Alan Turing Institute. He received the MS (Hons) degree in computer engineering from the University of Genoa, Genoa, Italy, in 2006, and the PhD degree in humanoid technologies from the University of Genoa and Italian Institute of Technology, in 2010. He was an Associate Researcher with the Takanishi Laboratory, Waseda University, Shinjuku, Japan, from 2010 to 2012, and with VisLab, Instituto Superior Técnico, Lisbon, Portugal, from 2013 to 2016. His main research interest is Cognitive Robotics, with more than 90 publications (h-index 25) on robotic manipulation, tactile sensing, sensorimotor learning and affordance perception.

Dimitri Ognibene is Associate Professor of Human Technology Interaction at the University of Milan-Bicocca, Italy. His main interest lies in understanding how social agents with limited sensory and computational resources can adapt to complex and uncertain environments, how this can induce suboptimal behaviors such as addiction or antisocial behaviors, and how this understanding can be applied to real-life problems. To this end, it develops both neural and Bayesian models and applies them both in the physical field, e.g. robots, and virtual settings, e.g. social media. Before joining Milano Bicocca University, he was Lecturer in Computer Science and Artificial Intelligence at the University of Essex, Marie Curie Actions COFUND fellow at University Pompeu Fabra (Barcelona, Spain), and Research Associate at the Centre for Robotics Research, Kings' College London, and previously at the Personal Robotics labs, Imperial College London and in the Institute of Cognitive Science and Technologies of the Italian Research Council (ISTC CNR). He is editor of *Cognitive Computation and Systems Journal*, handling editor of *Cognitive Processing*, review editor for *Paladyn, The Journal of Behavioral Robotics*, *Frontiers Bionics and Biomimetics*, and *Frontiers Computational Intelligence in Robotics*, guest associate editor for *Frontiers in Neurorobotics* and *Frontiers in Cognitive Neuroscience*. He has been chair of the robotics area of several conferences and workshops.

Takayuki Nagai received his BE, ME, and PhD degrees from the Department of Electrical Engineering, Keio University, in 1993, 1995, and 1997, respectively. Since 1998, he had been with the University of Electro-Communications, and from 2018 he has been a professor of the graduate school of Engineering Science, Osaka University. From 2002 to 2003, he was a visiting scholar at the Department of Electrical Computer Engineering, University of California, San Diego. He also serves as a specially appointed professor at UEC AIX, a visiting researcher at Tamagawa University Brain Science Institute, a visiting researcher at AIST AIRC. He has received IROS Best Paper Award Finalist, Advanced Robotics Best Paper Award, JSAI Best Paper Award, etc. His research interests include intelligent robotics, cognitive developmental robotics, and robot learning. He aims at realizing flexible and general intelligence like human by combining AI and robot technologies.

Alessandra Sciutti is Tenure Track Researcher, head of the CONTACT (COgNiTive Architecture for Collaborative Technologies) Unit of the Italian Institute of Technology (IIT). She received her BS and MS degrees in Bioengineering and the PhD in Humanoid Technologies from the University of Genova in 2010. After two research periods in USA and Japan, in 2018 she has been awarded the ERC Starting Grant wHiSPER (www.whisperproject.eu), focused on the investigation of joint perception between humans and robots. She published more than 80 papers and abstracts in international journals and conferences and participated in the coordination of the CODEFROR European IRSES project (https://www.codefror.eu/). She is currently Associate Editor for several journals, among which the *International Journal of Social Robotics*, the *IEEE Transactions on Cognitive and Developmental Systems* and *Cognitive System Research*. The scientific aim of her research is to investigate the sensory and motor mechanisms underlying mutual understanding in human-human and human-robot interaction.

Alvaro Costa-Garcia is a research scientist at the Intelligent Behavior Control Unit, CBS, RIKEN. He received MS degree in Telecommunications Engineering from the Systems Engineering and Automation Department at the Miguel Hernández University of Elche, Spain (2013). He obtained a PhD in the Program on Industrial and Telecommunications Technology from the same university in 2016. During the PhD, his research was focused on lower limb rehabilitation methods and the study of cognitive mechanisms related to gait through the evaluation of electroencephalographic signals. Currently, he is studying the neural processes behind voluntarily motor control through the evaluation of electromyography and its applications in the field of neuro rehabilitation.

Yohei Oseki is an Assistant Professor in the Department of Language and Information Sciences at the University of Tokyo and a visiting scholar at RIKEN Center for Advanced Intelligence Project (AIP). Before joining the University of Tokyo, he received a PhD from the Department of Linguistics at New York University in 2018 and was a visiting scholar at the Department of Linguistics at the University of Massachusetts Amherst and the Cold Spring Harbor Laboratory. His research integrates natural language processing with the cognitive and brain sciences of language and attempts to build machines that process natural language like humans. He founded Computational Psycholinguistics Tokyo (CPT) and also organizes Cognitive Modeling and Computational Linguistics (CMCL).

Tadahiro Taniguchi received his ME and PhD degrees from Kyoto University, in 2003 and 2006, respectively. From 2005 to 2008, he was a Japan Society for the Promotion of Science Research Fellow in the same university. From 2008 to 2010, he was an Assistant Professor at the Department of Human and Computer Intelligence, Ritsumeikan University. From 2010 to 2017, he was an Associate Professor in the same department. From 2015 to 2016, he was a Visiting Associate Professor at the Department of Electrical and Electronic Engineering, Imperial College London. Since 2017, he is a Professor at the Department of Information Science and Engineering, Ritsumeikan University, and a Visiting General Chief Scientist at the Technology Division of Panasonic Corporation. He has been engaged in research on machine learning, emergent systems, intelligent vehicles, and symbol emergence in robotics.

References

[1] Sinnett S, Smilek D, Kingstone A. Cognition. 6th ed. Ontario: Oxford University Press; 2016.
[2] Neisser U. Cognitive psychology. New York: Appleton-Century-Crofts; 1967.
[3] Byrne RW, Bates LA. Why are animals cognitive? Curr Biol. 2006;16:R445–R448.
[4] Lieto A, Bhatt M, Oltramari A, et al. The role of cognitive architectures in general artificial intelligence. Cogn Syst Res. 2006;48:1–3.
[5] Available from: https://www.ieee-ras.org/cognitive-robotics
[6] Available from: https://www.youtube.com/user/SSROVER77/playlists
[7] Brooks RA. A robust layered control system for a mobile robot. IEEE J Robot Automation. 1986;2:12–23.
[8] Nishihara J, Nakamura T, Nagai T. Online algorithm for robots to learn object concepts and language model. IEEE Trans Cogn Develop Syst. 2019;9(3):255–268.
[9] Aoki T, Nakamura T, Nagai T. Online joint learning of object concepts and language. IEEE ICDL-EPIROB, Workshop Second Workshop on Language Learning; 2017.
[10] Miyazawa K, Horii T, Aoki T, et al. Integrated cognitive architecture for robot learning of action and language. Frontiers in Robotics and AI. 2019;6:2280.
[11] Solak G, Jamone L. Learning by demonstration and robust control of dexterous in-hand robotic manipulation skills. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2019. p. 8246–8251.
[12] Kawamura A, Tahara K, Kurazume R, et al. Dexterous object manipulation by a multi-fingered robotic hand with visual-tactile fingertip sensors. J Robot Mechatron. 2013;25(1):125–135.
[13] Choi S-H, Tahara K. Robust visual servoing for object manipulation against temporary loss of sensory information using a multi-fingered hand-arm. ROBOMECH J. 2020;7(14):296.
[14] Andrychowicz M, Baker B, Chociej M, et al. Learning dexterous in-hand manipulation. Int J Rob Res. 2020;39(1):3–20.
[15] Falco P, Attawia A, Saveriano M, et al. On policy learning robust to irreversible events: an application to robotic in-hand manipulation. IEEE Robot Automat Lett. 2018;3(3):1482–1489.
[16] Okajima S, Tournier M, Alnajjar FS, et al. Generation of human-like movement from symbolized information. Front Neurorobot. 2018;12:436.
[17] Shimoda S, Yoshihara Y, Kimura H. Adaptability of tacit learning in bipedal locomotion. IEEE Trans Auton Ment Dev. 2013;5(2):152–161.
[18] Miyazaki F, Arimoto S. Singular perturbation for the analysis of biped locomotion system with many degrees of freedom. Trans Soc Instrument Control Eng. 1979;15(4): 498–504.
[19] Nakamura T, Nagai T, Taniguchi T. SERKET: an architecture for connecting stochastic models to realize a large-scale cognitive model. Preprint; 2017. Available from: arXiv:1712.00929.
[20] Rosenfeld A, Zemel R, Tsotsos JK. The elephant in the room; 2018.
[21] Marchegiani L, Newman P. Listening for sirens: Locating and classifying acoustic alarms in city scenes. Preprint; 2018. Available from: arXiv:1810.04989.
[22] Bajcsy R, Aloimonos Y, Tsotsos JK. Revisiting active perception. Auton Robots. 2020;42(2):177–196.
[23] Aloimonos J, Weiss I, Bandyopadhyay A. Active vision. Int J Comput Vision. 1988;1(4):333–356.
[24] Ballard DH. Animate vision. AI. 1991;48:57–86.
[25] Ognibene D, Baldassare G. Ecological active vision: four bioinspired principles to integrate bottom-up and adaptive top-down attention tested with a simple camera-arm robot. IEEE Trans Auton Ment Dev. 2014;7(1):3–25.
[26] Ramakrishnan SK, Jayaraman D, Grauman K. An exploration of embodied visual exploration. Int J Comput Vis. 2021;129:1616–1645.
[27] Tistarelli M, Sandini G. On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow. IEEE Trans Pattern Anal Mach Intell. 1993;15(4):401–410.
[28] Ognibene D, Chinellato E, Sarabia M, et al. Contextual action recognition and target localization with an active allocation of attention on a humanoid robot. Bioinspir Biomim. 2013;8(3):035002.
[29] Rothkopf CA, Ballard DH, Hayhoe MM. On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow. J Vis. 2007;7(14):1–20.

[30] Friston K, Rigoli F, Ognibene D, et al. Active inference and epistemic value. Cogn Neurosci. 2015;6(4):187–214.
[31] Nolfi S. Power and the limits of reactive agents. Neurocomputing. 2002;42:119–145.
[32] de Croon G, Postma EO, van den Herik HJ. A situated model for sensoryotor coordination in gaze control. Pattern Recognit Lett. 2006;27(11):1181–1190.
[33] de Croon GCHE, Sprinkhuizen-Kuyper IG, Postma EO. Comparing active vision models. Image Vis Comput. 2009;27(4):374–384.
[34] Andreopoulos A, Tsotsos JK. A computational learning theory of active object recognition under uncertainty. Int J Comput Vision. 2013;101(1):95–142.
[35] Kaelbling LP, Littman ML, Cassandra AR. Planning and acting in partially observable stochastic domains. Artif Intell. 1998;101(1):99–134.
[36] Tishby N, Polani D. Information theory of decisions and actions. New York (NY): Springer; 2011. 601–636.
[37] Shah D, Eysenbach B, Kahn G, et al. Ving: learning open-world navigation with visual goals. IEEE International Conference on Robotics and Automation; 2021.
[38] Bonet B, Geffner H. Belief tracking for planning with sensing: width, complexity and approximations. J Artificial Intelligence Res. 2014;50:923–970.
[39] Satsangi Y, Lim S, Whiteson S, et al. Maximizing information gain in partially observable environments via prediction rewards. Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems; 2020.
[40] Friston K, Adams R, Perrinet L, et al. Perceptions as hypotheses: saccades as experiments. Front Psychol. 2012;3:151.
[41] Krause A, Guestrin C. Near-optimal nonmyopic value of information in graphical models. Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence; 2005.
[42] Heckerman D, Horvitz E, Middleton B. An approximate nonmyopic computation for value of information. IEEE Trans Pattern Anal Mach Intell. 1993;15(3): 292–298.
[43] Ammirato P, Poirson P, Park E, et al. A dataset for developing and benchmarking active vision. In: 2017 IEEE International Conference on Robotics and Automation (ICRA); 2017. p. 1378–1385.
[44] Xia F, Zamir AR, Sax ZHA, et al. Gibson env: real-world perception for embodied agents. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018. p. 9068–9079.
[45] Geffner H. Model-free, model-based, and general intelligence. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18. International Joint Conferences on Artificial Intelligence Organization, 7; 2018. p. 10–17.
[46] Auer P, Cesa-Bianchi N, Fischer P. Finite-time analysis of the multiarmed bandit problem. Mach Learn. 2002;47(2):235–256.
[47] Audibert JY, Munos R, Szepesvári C. Exploration-exploitation tradeoff using variance estimates in multi-armed bandits. Theor Comput Sci. 2009;410(19): 1876–1902.
[48] Jayaraman D, Grauman K. End-to-end policy learning for active visual categorization. IEEE Trans Pattern Anal Mach Intell. 2019;41(7):1601–1614.
[49] Whitehead SD, Lin L-J. Reinforcement learning of non-markov decision processes. Artif Intell. 1995;73(1): 271–306.
[50] Paletta L, Pinz A. Active object recognition by view integration and reinforcement learning. Rob Auton Syst. 2000;31(1):71–86.
[51] Scete M, Doyle J. Bow ties, metabolism and disease. Trends Biotechnol. 2004;22(9):446–450.
[52] Ting LH. Dimensional reduction in sensorimotor systems: a framework for understanding muscle coordination of posture. Prog Brain Res. 2007;165:299–321.
[53] Alnajjar FSK, Itkonen M, Berenz V, et al. Sensor synergy as environmental input integration. Frontier in Neuroscience. 2015;15:436.
[54] D'Avella A, Saltiel P, Bizzi E. Combinations of muscle synergies in the construction of a natural motor behavior. Nat Neurosci. 2003;6:300–308.
[55] Ivanenko YP, Poppele RE, Lacquaniti F. Spinal cord maps of spatiotemporal alpha-motoneuron activation in humans walking at different speeds. J Neurophysiol. 2006;95(2):602–618.
[56] Yang N, An Q, Kogami H, et al. Temporal features of muscle synergies in sit-to-stand motion reflect the motor impairment of post-stroke patients. IEEE Trans Neural Syst Rehabil Eng. 2019;27(10):2118–2127.
[57] Okajima S, Alnajjar FS, Costa A, et al. Theoretical approach for designing the rehabilitation robot controller. Adv Robot. 2019;33(14):674–686.
[58] Yamasaki H, Itkonen M, Alnajjar F, et al. Model for prompt and effective classification of motion recovery after stroke considering muscle strength and coordination factors. J Neuroeng Rehabil. 2019;16(1):1–11.
[59] Sciutti A, Patane L, Nori F, et al. Understanding object weight from human and humanoid lifting actions. IEEE Trans Auton Ment Dev. 2014;6(2):80–92.
[60] Sandini G, Mohan V, Sciutti A, et al. Social cognition for human-robot symbiosis-challenges and building blocks. Front Neurorobot. 2018;12:34.
[61] Wilson M. Six views of embodied cognition. Psychon Bull Rev. 2002;9:625–636.
[62] Pfeifer R, Lungarella M, Sporns O. 7 – the synthetic approach to embodied cognition: a primer. In: Calvo P, Gomila A, editors. Handbook of cognitive science. San Diego (CA): Elsevier; 2008. p. 121–137. (Perspectives on Cognitive Science).
[63] Vernon D, Lowe R, Thill S, et al. Embodied cognition and circular causality: on the role of constitutive autonomy in the reciprocal coupling of perception and action. Front Psychol. 2015;6:1660.
[64] Jamone L, Ugur E, Cangelosi A, et al. Affordances in psychology, neuroscience and robotics: a survey. IEEE Trans Cogn Develop Syst. 2018;10(1):4–25.
[65] Bassili JN. Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face. J Pers Soc Psychol. 1979;37(11):2049–2058.
[66] Ueda S, Kitaoka A, Suga T. Wobbling appearance of a face induced by doubled parts. Perception. 2011;40(6): 751–756.

[67] Scherer KR. Expression of emotion in voice and music. J Voice. 1995;9(3):235–248.
[68] Karg M, Samadani A-A, Gorbet R, et al. Body movements for affective expression: a survey of automatic recognition and generation. IEEE Trans Affect Comput. 2013;4(4):341–359.
[69] Laeng B, Sirois S, Gredebäck G. Pupillometry: a window to the preconscious? Perspect Psychol Sci. 2012;7(1): 18–27.
[70] Flanagan JR, Johansson RS. Action plans used in action observation. Nature. 2003;424(6950):769–771.
[71] Knoblich G, Butterfill S, Sebanz N. Chapter three – psychological research on joint action: theory and data. Psychol Learn Motiv. 2011;54:59–101.
[72] Noceti N, Sciutti A, Rea F. Modelling human motion. New York City: Springer; 2020.
[73] Sandini G, Sciutti A, Rea F. Movement-based communication for humanoid-human interaction. New York City: Springer; 2019. p. 2169–2197.
[74] Sciutti A, Mara M, Tagliasco V, et al. Humanizing human-robot interaction: on the importance of mutual understanding. IEEE Technol Soc Mag. 2018;37(1): 22–29.
[75] Haggard P, Eitam B. The sense of agency. Oxford: Oxford University Press; 2015.
[76] Evans V. Language and time: a cognitive linguistics approach. Cambridge: Cambridge University Press; 2013.
[77] Ueda S, Shimoda S. Subjective time compression induced by continuous action. Sci Rep. 2021;11:1–11.
[78] Ueda S, Nakashima R, Kumada T. Influence of levels of automation on the sense of agency during continuous action. Sci Rep. 2021;11:1–11.
[79] Legaspi R, Toyoizumi T. A bayesian psychophysics model of sense of agency. Nature Commun. 2019;10(1): 1–11.
[80] Mazzola C, Mois Aroyo A, Rea F, et al. Interacting with a social robot affects visual perception of space. In: Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction. New York (NY): Association for Computing Machinery; 2020. p. 549–557.
[81] Di Cesare G, Vannucci F, Rea F, et al. How attitudes generated by humanoid robots shape human brain activity. Sci Rep. 2020;10(1):1–12.
[82] Lake BM, Ullman TD, Tenenbaum JB, et al. Building machines that learn and think like people. Behav Brain Sci. 2017;40:e253.
[83] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: Proceedings of NIPS; 2017. p. 5998–6008.
[84] Kuribayashi T, Oseki Y, Ito T, et al. Lower perplexity is not always human-like. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics; 2021.
[85] Dyer C, Kuncoro A, Ballesteros M, et al. Recurrent neural network grammars. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego (CA): Association for Computational Linguistics; Jun 2016. p. 199–209.
[86] Noji H, Oseki Y. Effective batching for recurrent neural network grammars. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics; 2021.
[87] Marr D. Vision: a computational investigation into the human representation and processing of visual information. Cambridge (MA): MIT Press; 1982.
[88] Kriegeskorte N, Douglas P. Cognitive computational neuroscience. Nat Neurosci. 2018;21:1148–1160.
[89] Taniguchi A, Hagiwara Y, Taniguchi T, et al. Spatial concept-based navigation with human speech instructions via probabilistic inference on Bayesian generative model. Adv Robot. Sept 2020;34(19):1213–1228.
[90] Newell A. Physical symbol systems. Cogn Sci. 1980;4(2): 135–183.
[91] Taniguchi T, Ugur E, Hoffmann M, et al. Symbol emergence in cognitive developmental systems: a survey. IEEE Transactions on Cognitive and Developmental Systems; 2018.
[92] Tangiuchi T, Mochihashi D, Nagai T, et al. Survey on frontiers of language and robotics. Adv Robot. 2019;33(15–16):700–730.
[93] Taniguchi A, Hagiwara Y, Taniguchi T, et al. Online spatial concept and lexical acquisition with simultaneous localization and mapping. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE; 2017. p. 811–818.
[94] Taniguchi A, Hagiwara Y, Taniguchi T, et al. Improved and scalable online learning of spatial concepts and language models with mapping. Auton Robots. 2020;44:927–946.
[95] Levine S. Reinforcement learning and control as probabilistic inference: tutorial and review. Report; 2018.
[96] Taniguchi T, Nakamura T, Suzuki M, et al. Neuro-SERKET: development of integrative cognitive system through the composition of deep probabilistic generative models. New Generation Comput. 2020;38:23–48.
[97] Taniguchi T, Yamakawa H, Nagai T, et al. Whole brain probabilistic generative model toward realizing cognitive architecture for developmental robots; 2021.