18

Learning Action-Perception Cycles in Robotics

A Question of Representations and Embodiment

Jeannette Bohg and Danica Kragic

Abstract

Since the 1950s, robotics research has sought to build a general-purpose agent capable of autonomous, open-ended interaction with realistic, unconstrained environments. Cognition is perceived to be at the core of this process, yet understan#ding has been challenged because cognition is referred to differently within and across research areas, and is not clearly defined. The classic robotics approach is decomposition into functional modules which perform planning, reasoning, and problem solving or provide input to these mechanisms. Although advancements have been made and numerous success stories reported in specific niches, this systems-engineering approach has not succeeded in building such a cognitive agent.

The emergence of an action-oriented paradigm offers a new approach: action and perception are no longer separable into functional modules but must be considered in a complete loop. This chapter reviews work on different mechanisms for action-perception learning and discusses the role of embodiment in the design of the underlying representations and learning. It discusses the evaluation of agents and suggests the development of a new *embodied* Turing Test. Appropriate scenarios need to be devised in addition to current competitions, so that abilities can be tested over long time periods.

Introduction

In June 2014, the University of Reading reported that a machine passed the famous Turing Test: a computer program impersonated a 13-year-old Ukrainian boy, called Eugene Goostman, and was able, through text interface, to make a

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science," Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18, series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

sufficient number of interrogators believe that they were communicating with an actual human being. This news attracted quite a bit of attention but not as much as one might have thought, since passing the Turing Test had been perceived to be proof that machines could think.

Turing proposed a general test of intelligence to measure the competency of an artificial system "in all purely intellectual fields." He believed that by the year 2000, machines would be capable of this mental process, classically labeled cognition. He discussed the problems associated with deciding when a machine would convincingly reach this level and proposed that the ambiguous question of whether machines could think be replaced by an imitation game which the machine would have to win to prove cognitive competency (Turing 1950). This test, he imagined, would assess the intellectual capabilities of the agent *independent* of the actual mechanism or principle behind it. Turing's original proposal and subsequent versions of the test (e.g., as is used in the Loebner Prize) did not attract significant attention in the robotics community.

One possible explanation for this relative disinterest might be found in an interesting parallel (Russell and Norvig 2003) between the quest for artificial intelligence (AI) and artificial flight: Aeronautical engineering is not defined as making "machines that fly so exactly like pigeons that they can fool even other pigeons." Aeronautic researchers are interested in the principles of aerodynamics. Thus, by analogy, AI researchers seem to have been interested in uncovering the underlying principles of intelligence rather than in duplicating an exemplar.

Early Approaches in Artificial Intelligence

In addition to suggesting the test, Turing (1950) theorized about the underlying principles. He favored the idea of a learning machine whose brain would be similar to that of a child (i.e., a blank slate). Certain built-in rules of operation for logical inference were possible, but these would be subject to change during learning. Interestingly, however, he did not consider it necessary for the agent to have limbs or eyes.

Subsequent researchers have dedicated significant attention to the problems that a machine would face during the proposed imitation game: understanding and producing natural language text, representing general knowledge and information from the ongoing conversation, reasoning to answer questions or to draw novel conclusions, and learning from experience to adapt to new situations. Special symbolic rule-based planners were developed that rely on the existence of an internal world model. Given a certain world state and a goal, these planners could devise a strategy to attain this goal. The first planning system, STRIPS, was developed for the Shakey Robot at Stanford Research Institute (Fikes and Nilsson 1971) and functioned independently to how the robot built the world model, recognized certain objects, or executed planned

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

actions. These problems were supposed to be solved independently by general-purpose, task-independent black-box modules. Significant progress was made early on in terms of these planning algorithms. Based on this work, supercomputers are now able to beat the best human players in chess or Jeopardy. Yet robots are still unable to demonstrate the autonomy and skill of a one-year-old child in terms of perception and motor control. We suggest that the greatest challenge to developing a general-purpose autonomous agent arises at the interface between the agent and the world, not at logical reasoning over readily given abstract symbols.

Emergence of the Action-Oriented Paradigm in Robotics

Moravec (1988) pointed out the following paradox: high-level symbolic reasoning, which requires relatively high effort by humans, seems to be relatively easy to automatize. However, tasks that humans can perform effortlessly (e.g., grasping of arbitrary objects or manipulation of tools) seem difficult for machines to achieve. While we are consciously aware of symbolic reasoning, these latter tasks are controlled by subconscious processes and thus they are much harder to reproduce. Moravec claims that these processes developed over thousands of years of evolution while abstract reasoning is a rather recent development.

Related to this, Brooks (1990, 1991b) proposed a new way to think about artificial intelligence. In contrast to Turing, he believed that a machine needs limbs and eyes to interact with a complex and dynamic environment. He rejected the focus on internal general-purpose representations of the world, symbolic reasoning, and a functional decomposition of intelligence. Instead, he defined intelligence in terms of a combination of simple behaviors, which were defined by directly connecting perception modules to controllers. These behaviors were combined in a structure that Brooks called the *subsumption architecture*. He could show that robots using this idea would expose intelligent behavior in dynamic and cluttered environments. They were even able to show simple grasps and navigate in mapped environments, without any need for complex internal representation and reasoning. These robots had no memory; instead they relied on sensors for continuous feedback from the world around them.

Parallels between Cognitive Science and Robotics

Almost simultaneously to Brooks' proposal, the early work of Varela et al. (1992) established the "enactive approach" to cognition. Similarly to Brooks, Varela and colleagues did not consider cognition to be the process of extracting general-purpose, task-independent representations of the world. Instead, they held that cognitive processes of internalizing the external and building structures are guided by action. This is further related to the internal simulation

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

theory in which the brain simulates the environment and reasons on it before acting (Jeannerod 1988). Clark's action-oriented representation (Clark 1998) and O'Regan and Noë's (2001) sensorimotor contingency (SMC) theory both support this work. According to SMC theory, the agent's SMCs are constitutive for cognitive processes and are defined as law-like relations between movements and associated changes in sensory inputs that are produced by the agent's actions. Accordingly, "seeing" cannot be understood as the processing of an internal visual "representation"; seeing corresponds to being engaged in a visual exploratory activity, mediated by knowledge of SMCs. Additional evidence from psychology and neuroscience stipulates that action in biological systems participates as a generative model in perceptual processes and in structuring knowledge about the world (Gallese et al. 1996; Fadiga et al. 1999; Borroni et al. 2005).

Robotics Research Today

Although the proposal by Brooks is now widely considered to have marked a paradigm shift in robotics, it still remains to be shown whether these ideas can yield more high-level autonomous behavior than that of insects. However, the robotics community has placed more research effort on the interface between an agent and its environment. This does not mean that robotics has agreed on one approach. The classic approach to AI and new directions coexist and are potentially combined. As it happens, this situation is similar to the development that occurred in cognitive science (Engel et al. 2013).

Current research in robotics is largely shaped by the *systems engineering approach* (Brock 2011), which very often aims at solving problems related to a specific application. Within the current research funding landscape, progress has to be fast and verifiable. Today's robots are complex systems that require expertise in many different subjects. A roboticist may commonly be specialized in one of them and try to abstract away the others. For example, researchers who are experts in control may know little about visual perception. Therefore, these modules are abstracted away and treated as black boxes. Any potentially complex two-sided interaction between control and perception is replaced by a simplified interface. Representations may be treated as general purpose and task independent. Very often, symbolic planners sit at the center of these approaches and devise plans that are computed over symbols provided by the black-box perception modules. Resulting action sequences are often executed in an open-loop manner without checking to see whether the expected effect has actually been achieved.

The research area of computer vision is rooted in the demand of robotics research for general-purpose, task-independent representations of semantic entities (Horn 1986). Due to the difficulties inherent in this problem, research in computer vision has developed away from robotics and now has little to do with it. First and foremost, it rarely addresses challenges that arise in robotics,

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

such as real-time requirements or the possibility to act in the environment for exploration. Similarly, in the area of control, the generation of movement is mainly studied in isolation. Feedback controllers usually close the loop around joint angles, velocities, or motor torques. Not as much focus has been placed on feedback about the environment structure. If this kind of feedback is required, it is often provided by precise motion capture systems in the hope that sometime in the future computer vision researchers will deliver the promised reliable general-purpose black boxes.

These approaches have brought tremendous progress in their associated research areas. However, when trying to unite them within a robotics system through the systems-engineering approach, they are only successful in restricted application scenarios that are not open-ended, largely static, and controlled. In general, current robots lag surprisingly far behind humans although they have faster and less noisy sensors and actuators and can perform rapid decision making and control (Wolpert et al. 2011). If we are still striving to discover the underlying fundamental principles between autonomous and purposeful behavior, we have not yet found the key.

Many people believe that the action-oriented paradigm offers the key to permit new insights. We have seen the emergence of the field of developmental robotics, which strives toward learning machines, already proposed by Turing. However, developmental roboticists also emphasize the importance of the learning agent being embodied (Lungarella et al. 2003). Below, we review a portion of the work that follows the action-oriented paradigm and focus mainly on mechanisms for action-perception learning. We focus on the role of embodiment in the design of the underlying representations as well as for the specific learning mechanism.

Representations

General-purpose autonomous robots cannot be preprogrammed for all the tasks they will be required to do; just like humans, they should be able to gather information from different sources and learn from the experiences of both humans and other robots. Thus, the ability to acquire new skills and adapt existing ones to novel tasks and contexts is a necessity. For some natural domains (e.g., cooking and meal preparation), models or plans for different tasks are already available. For example, web pages such as ehow.com and wikihow.com provide simple and detailed instructions on how to plant a tree or make lemon curd. These sites contain thousands of directives for everyday activities: about 45,000 on wikihow.com and more than 250,000 on ehow.com. Using written and structured instructions is common for humans. Many repetitive and dangerous tasks in factories and laboratories have natural language and graphic workflow specifications that are similar to task instructions in the World Wide Web.

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

A natural idea is to enable robots to do the same. However, this poses several challenges. To look at, listen to, and perceive an instruction, robots need to be able to understand text, video, spoken commands, or even all of them at the same time. They need to be able to understand concepts that are symbolic and relate them to sensory information. For example, an object such as "fork" in spoken or written instructions needs to relate to specific visual features that can be extracted from an image. In addition, the representation of a fork needs to be such that the robot can distinguish it from a knife.

A classic approach is to develop general-purpose, task-independent representations of semantic entities needed in these aforementioned tasks. Representations like this promise effectiveness through compression of a lot of information into a single symbol, which then is able to generalize to all possible situations and contexts.

Humans use visual and other sensory feedback extensively to plan and execute actions. However, this process is not a well-defined one-way stream: how we plan and execute actions depends on what we already know about (a) the environment in which we operate (context), (b) the action we are about to undertake (task), and (c) the result expected from our actions (effect). This insight has been picked up in robotics and resulted in many models that try to represent actions and percepts jointly instead of finding *the one* representation that matches all purposes. The concept of affordances, as proposed by Gibson (1977), has inspired representations, especially in grasping and interaction with objects.

To a certain degree, affordances can be observed in images. In several works (Bohg and Kragic 2009; Saxena et al. 2008; Stark et al. 2008), relations between visual cues and grasping affordances are learned from training data. In Stark et al. (2008), object grasping areas are extracted from short videos of humans interacting with the objects. In Bohg and Kragic (2009) as well as Saxena et al. (2008), a large set of two-dimensional object views are labeled with grasping points. Early work on functional object recognition (Rivlin et al. 1995; Stark and Bowyer 1996) can be seen as a first step toward recognizing affordances from images. Objects are modeled in terms of their functional parts (e.g., handle, hammerhead; Rivlin et al. 1995) or by reasoning about shape in association to function (Stark and Bowyer 1996). In these approaches, the relation between objects and action is usually predefined by humans.

As pointed out by Sloman (2001), we are not consciously aware of a significant amount of human visual processing: we do not experience using optical flow to control our posture nor are we aware of the saccades and fixational eye movements that allow us to negotiate with the complexity of everyday scenes (Koch and Ullman 1985). Therefore, these processes are not easy for us to reproduce in an artificial system. It has been argued that representations should only be constructed by the system itself through interaction with and exploration of the world rather than through *a priori* specification or programming (Granlund 1999). Thus, objects should be represented as invariant

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

combinations of percepts and responses, where the invariances (which are not restricted to geometric properties) need to be learned through interaction rather than specified or programmed *a priori* (Granlund 1999). A system's ability to interpret the external world is dependent on its ability to interact with it. This interaction structures the relationship between perception and action.

In robotics, this can be a slow process, due to the challenges involved when extensive physical interaction is required. Over the last several years, however, advanced oculomotor and hand-eye systems have been demonstrated (e.g., Moren et al. 2008; Montesano et al. 2008; Kraft et al. 2008; Rasolzadeh et al. 2010). There are approaches that let the robot interact with its environment and learn through trial and error. One example is a cognitive model for grasp learning in infants (Oztop et al. 2005). A model for learning affordances using Bayesian networks embedded within a general developmental architecture has been proposed by Montesano et al. (2008). Kraft et al. (2008) proposed object-action complexes as semi-supervised procedures for encoding sensorimotor relations and showed how this can be used to improve the robot's inner model and behavior. The idea was further developed by Song et al. (2010), where the relationships between object, action, constraint features, and task were encoded using Bayesian networks.

These approaches often consider actions at discrete moments in time (e.g., a grasp when approaching but not yet making contact with the object) or as a discrete symbol (e.g., pushing, pulling, grasping, pouncing). The representation of movement over time and how to couple it to sensory input is also an active area of research. One popular representation of this kind has been dynamic movement primitives (Ijspeert et al. 2002; Schaal et al. 2007), proposed for both feedforward and feedback motor commands. Dynamic movement primitives relate to optimal control theory approaches such as minimum jerk trajectories (Flash and Hogan 1985), as well as machine learning approaches such as hidden Markov models (Billard et al. 2004; Inamura et al. 2004). Dynamic movement primitives have been coupled to sensory input through maintaining a visual representation of the goal point and adapting the goal's tracked position (Pastor et al. 2009). Other ways to shortcut perception have been to use motion capture systems or easy-to-detect fiducial markers (Calinon 2009). Only recently have we seen how low-level sensory feedback can define the goal directly (Pastor et al. 2011). During execution, a dynamic movement primitive is adapted such that the robot *feels* the same as when the movement was demonstrated.

Sensorimotor knowledge in humans is structured as of childhood, and this type of lifelong learning has been the focus of developmental approaches in robotics (Pfeifer and Scheier 1999; Kuniyoshi et al. 2003; Lungarella et al. 2003). Thus far, however, developmental approaches have been demonstrated on rather simplistic problems. If a large corpus of data is available, informed learning approaches, such as imitation learning (Schaal 1999), can be applied.

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

Learning and Priors

Much of the work on learning and priors was inspired by Piaget's ideas of assimilation and accommodation. These two complementary processes of adaptation enable the experiences of the external world to be internalized. The problem of assimilation has been addressed more widely, given that some predefined structure has been used for classification of new experiences. The problem of accommodation requires the representation of knowledge structure to be changed as the new data is gathered and requires more advanced learning techniques to be employed.

An organism cannot develop without some built-in ability. However, if all abilities are built in, the organism is unable to develop. There is an optimal level for how much phylogeny should provide versus how much needs to be acquired during the lifetime.

A human spends years interacting with its environment before it can master certain complex cognitive or motor tasks. At the same time, robots and the computational modules are often expected to learn from very little data. Imitation may be a very good way to bootstrap an artificial system. Even then, during its "lifetime," a robot will encounter so many more situations than what could possibly have been demonstrated to it by a human teacher. In fields such as computer vision or speech processing, "big data" (visual or auditory data annotated with strong or weak semantic labels) has become increasingly more available, impacting the very type of research that is being performed in these areas. Methods that can be trained on these massive amounts of data are currently outperforming previous state-of-the-art approaches (Halevy et al. 2009). In robotics, there are no labeled databases of this order of magnitude to help bootstrap the system. This is most likely due to the complexity of a robot system, which makes it hard to collect and label these massive amounts of data. In contrast to computer vision and speech-processing data, a data point in a robotics database also depends on an action. Therefore, the usual assumption of independent and identically distributed data cannot be as easily made in a robotics system.

Robotic systems receive a continuous stream of sensory (visual, haptic, auditory) data that is currently largely unused. Data is often extracted at arbitrary discrete points in time and then processed independently of the other data points in the time series. Exceptions to this are feedback controllers that enable robots to execute movements toward a goal or along a trajectory. Feedback on some state variable that is actively controlled is continuously gathered, and the appropriate action is computed to minimize the error between the actual and desired state. Most commonly, these state variables contain joint angles and velocities, forces, and torques which act on joints directly or on end-effectors. The state may also contain information from vision sensors, such as the pose of objects in the environment. For these quantities, good models (e.g., rigid body motion or dynamics) exist to help the controllers design and compute the next

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

best action based on carefully selected sensory feedback. For more complex or multimodal data or more complex goals, modeling the mapping between the sensory state and next best action becomes much harder.

Although the "big data" paradigm holds great promise for learning some of these aspects, it is unclear how the data which a robotics system produces (time-series, multimodal and synchronized in time, structured) can be leveraged. Some examples exist for learning from time-series data, but the majority of work focuses on learning from discrete data samples.

Currently, the big data paradigm considers the problem of discovering *correlations* in data. However, robotics seems to be largely dominated by another structuring principle in data: causality. Discovering causality from data is difficult (Pearl 2009). Nonetheless, intuitively, understanding causality seems to be the key to predict the changes in sensory percepts after an agent executes an action.

Embodiment and Imitation

From the viewpoint of morphology, our bodies, actuators, and sensors exist to support effective action (Kuypers 1973) but there is nothing from the perspective of robotic systems that requires a cognitive system to take human shape. Ziemke's framework of embodied systems distinguishes five types of embodiment: structural coupling, historical embodiment, physical embodiment, organismoid embodiment, and organismic embodiment (Ziemke 2003). A single type of embodiment, however, cannot guarantee that the resultant cognitive behavior will be in any way consistent with human models or concepts.

Transfer of information between a teacher (human/robot) and a student (robot) requires a common knowledge representation. When the human and student have identical motor and sensory capabilities, the task may be simply to transform the action of one to the other by changing the frame of reference. Such transfers are not commonly possible, given that embodiments and associated capabilities often differ. To ensure compatibility with human concepts, there may be a need for higher similarity to humans regarding physical movement, interaction, exploration, and perhaps even human form (Brooks 2002).

In terms of object grasping and manipulation, the naïve approach to facilitate grasp transfer between different embodiments is to model the observed action of the teacher and map all the action parameters to the robot hand, which is commonly referred to as the "action-level" imitation (Alissandrakis et al. 2002). However, since different embodiments have different capabilities, the action required to achieve the goal may be different.

Imitation learning is an effective approach for teaching robots simple tasks (Billard et al. 2008). The learning paradigm based on an internal model (Wolpert and Kawato 1998) has received considerable attention. The work by Rao et al. (2007) implements an internal model through Bayesian networks.

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

Demiris and Johnson (2003) show that the internal models that represent the brain circuitry subserving sensorimotor control also participate in action recognition. They are used to predict the goal of observed behavior and activate the correct actions to maintain or achieve the "goal" state. Later work (Oztop et al. 2005) extends the use of an internal model to the domain of visual-manual tasks. We believe that future research in this area will address the interplay between the embodiment, knowledge representation, and learning in more detail. Abstraction from the embodiment may be a key, but one wonders to what extent this is reasonable to do.

Evaluation and Verification: An Embodied Turing Test

While the aforementioned proposals have been verified in specific scenarios and applications, we lack an understanding of how big their potential is toward the development of general-purpose cognitive agents capable of autonomous and open-ended interaction with realistic, unconstrained environments. How can an action-oriented approach actually be verified and compared to other approaches?

Several tests have been proposed to evaluate general cognitive capabilities of an artificial agent, the most famous being the Turing Test. Turing (1950) was interested in the potential mechanisms and principles behind rational human reasoning, which we nowadays summarize with the somewhat fuzzy term of cognition. Instead of evaluating these mechanisms themselves, he proposed to measure the resemblance of an agent to a real person in a dialogue scenario. In this way, he proposed a way to circumvent the difficult problem of defining precisely the mental process of *thinking*. Turing believed that the exact computational structure of the mechanism does not matter as long as the artificial agent is perceived to perform rational human reasoning. Furthermore, he believed that equipping the machine with a body was entirely beside the point. The actual Turing Test has played a significant role in the field of human-machine interaction. Although new versions of this test have been proposed (Harnad 1991; Marcus 2014), only the total Turing Test begins to test sensorimotor capabilities.

To verify the methods proposed by the action-oriented paradigm, do we need a new *embodied* Turing Test?

One option would be to take the Turing Test and use it to evaluate not the actual mechanisms but rather the resemblance of how an artificial agent acts in the world compared to how a person would act. An external observer would decide whether a robot that is performing certain tasks in an environment is acting autonomously or is teleoperated. The tasks for the robot could involve manipulation and locomotion tasks of different degrees of difficulty, in different environments (e.g., household or disaster relief scenarios). Tasks could also involve physical interaction or collaboration with other agents or humans (e.g.,

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

preparing a meal, clearing a dinner table, assembling furniture, rescuing a person from a disaster site, or collaborating with a person to perform assembly tasks).

The advantage of this type of test is that a specific task would need to be autonomously performed in a fluid manner that resembles how a human would perform the task. We imagine this human *manner* to involve a certain level of dexterity, flexibility in the presence of a dynamically changing environment, as well as robustness to noise and failures.

It is debatable whether the criterion of resemblance to human fluidity when performing a task is desirable. It may be important in tasks where the robot is collaborating or interacting with humans such that its actions are predictable, but may have limited relevance when it comes to other (e.g., household) tasks.

A Robotics Challenge

Several robotics competitions have been set up to evaluate the performance of artificial agents, not only for purely intellectual tasks but also for tasks involving physical interaction: RoboCup, RoboCup@Home, DARPA Learning Locomotion, DARPA Autonomous Robotic Manipulation, and the DARPA Robotics Challenge. These challenges usually have well-defined goals that revolve around a specific scenario, such as soccer, and can easily be verified. The scenarios are usually formulated broadly so that the goals can be adapted and made more or less difficult from phase to phase. Furthermore, such competitions have the ability to bundle forces and focus them onto one goal (Marcus 2014). The spirit of competition seems to be a powerful source of motivation among researchers.

Do We Need a New Embodied Turing Test?

Without a doubt, it would be advantageous to have a test that could easily evaluate a set of well-defined goals. However, defining these goals poses the initial challenge. Although competitions can serve as a powerful motivator, experience shows that no matter how carefully such goals are defined or out of which original question they came, what counts in the end is winning. The hope or intention of discovering principles behind, for example, intelligence or autonomy, may be rejected in favor of *getting the task done*. Certainly this can be observed in earlier attempts to pass the Turing Test through the use of parlor tricks and purposeful deceit (Marcus 2014). All of the above-mentioned robot competitions encourage what is commonly referred to as *hacking*; that is, engineering solutions which exploit a fixed structure in a scenario, thus sacrificing the generality of solutions. Nevertheless, competitions do offer clear demonstrations of which type of task can be achieved, and some fundamental insights are inevitably gained, although less than what one would hope for or expect.

Once a problem has been solved and its inner workings computationally and algorithmically revealed, we often no longer believe that the associated

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

artificial system is intelligent. It is *just* computation. We wonder whether it is appropriate to let a person judge the resemblance of an agent executing a task to a real human doing the same. This may shift the focus from fulfilling a specific task to that of doing it robustly and fluidly.

Conclusions

Much can be said about perception and action as well as the work that has been done over the last sixty years. Here we reviewed cases which show some of the relations across different fields of research. Importantly, cognition is a process that needs to be studied and approached as such. Proper representations and learning mechanisms are necessary to meet the goal of developing autonomous agents capable of open-ended interaction. Equally important is the issue of how to assess and verify that an agent has made the proper choices. To evaluate and verify the capabilities of a robot, we suggest that the well-known Turing Test be reworked into an *embodied* Turing Test. Many scenarios can be envisioned for such a test; however, we believe short, competition-like scenarios are insufficient. Appropriate scenarios need to be devised that will test a robot's ability over long periods of time.

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.