Metrics and benchmarks in human-robot interaction: Recent advances in cognitive robotics

Amir Aly, Sascha Griffiths, Francesca Stramandinoli

► **To cite this version:**

HAL Id: hal-01507342

https://hal.archives-ouvertes.fr/hal-01507342

Submitted on 13 Apr 2017

Metrics and benchmarks in human-robot interaction: Recent advances in cognitive robotics

Amir Aly [a,*], Sascha Griffiths [b], Francesca Stramandinoli [c,1]

[a] *University of Paris-Saclay, ENSTA ParisTech, Palaiseau, France*
[b] *Queen Mary University, London, England, United Kingdom*
[c] *Italian Institute of Technology (IIT),Genoa, Italy*

Abstract

Robots are having an important growing role in human social life, which requires them to be able to behave appropriately to the context of interaction so as to create a successful long-term human-robot relationship. A major challenge in developing intelligent systems, which could enhance the interactive abilities of robots, is defining clear metrics and benchmarks for the different aspects of human-robot interaction, like human and robot skills and performances, which could facilitate comparing between systems and avoid application-biased evaluations based on particular measures. The point of evaluating robotic systems through metrics and benchmarks, in addition to some recent frameworks and technologies that could endow robots with advanced cognitive and communicative abilities, are discussed in this technical report that covers the outcome of our recent workshop on current advances in cognitive robotics: *Towards Intelligent Social Robots - Current Advances in Cognitive Robotics*, in conjunction with the 15th IEEE-RAS Humanoids Conference - Seoul - South Korea - 2015 (https://intelligent-robots-ws.ensta-paristech.fr/). Additionally, a summary of an interactive discussion session between the workshop participants and the invited speakers about different issues related to cognitive robotics research is reported.

1. Introduction and related work

The fast emerging interdisciplinary research area of cognitive robotics focuses on developing intelligent robots that can perceive the environment, act, and learn from experience so as to adapt their generated behaviors to interaction

* Corresponding author.

E-mail addresses: amir.aly@ensta-paristech.fr (A. Aly), sascha.griffiths@qmul.ac.uk (S. Griffiths), francesca.stramandinoli@iit.it (F. Stramandinoli).

in an appropriate manner. The growing collaboration between different disciplines within cognitive robotics, in addition to the accelerating advances in robotics related technologies, took robots a step closer to fill the gap separating theoretical research and real world applications. This requires defining evident criteria for evaluating robot-based applications in our life. While several theoretical studies in the related literature focused on setting groups of metrics and benchmarks in cognitive robotics so as to correlate between findings and compare between systems, we present a subjective evaluation for some common metrics, benchmarks, and frameworks through a Likert scale questionnaire, in which all questions are presented on a 7-point scale.

Steinfeld et al. (2006) proposed a set of metrics covering different task groups, such as: navigation (e.g., mobile robot navigation in the environment so as to manage path planning, available resources, avoiding obstacles, etc.), management (i.e., coordinating the actions of robots and human users), manipulation (i.e., physical robot interaction with the environment), and social (i.e., social interaction tasks between robots and human users). Similarly, Olsen and Goodrich (2003) offered a group of metrics for evaluating human-robot interaction including: task effectiveness (i.e., efficiency of a human-robot team in performing a task), neglect tolerance (i.e., robot performance drop when human attention is transferred to other tasks), robot attention demand (i.e., level of required human attention for instructing a robot), fan out (i.e., number of robots that a human user can operate at a time), and interaction effort (i.e., required time and cognitive potential of a human user to interact with a robot). Hoffman (2013) designed a group of subjective and objective metrics for evaluating fluency in human-robot collaboration. The subjective metrics of fluency include: human trust in robot, robot contribution to team, and human-robot team performance improvement over time. Meanwhile, the objective metrics of fluency include: robot idle time (i.e., robot waiting time for additional inputs from the human in order to take a decision), human idle time (i.e., human waiting time for the robot to complete finishing an assigned task), concurrent activity (i.e., active time of both the robot and the human), and functional delay (i.e., time difference between the end of one agent's task, whether the human or the robot, and the beginning of the other agent's task).

Moreover, Kahn et al. (2007) proposed a group of psychological benchmarks in order to evaluate human-like robots empirically. These benchmarks consider the basic characteristics of human behaviors - so as to increase the naturalness of robot behaviors as much as possible - including robot autonomy and imitation of people's behaviors, in addition to benchmarks focusing on the moral aspects of robot behaviors and their consequences on interaction with human users. Similarly, Rentschler, Cooper, Blasch, and Boninger (2003) and Feil-Seifer, Skinner, and Mataric (2007) addressed safety (i.e., level of robot safety for its human users during interaction with the environment) and scalability (i.e., ability of the robot system to be used by different user populations) as important benchmarks for robot system evaluation in socially assistive robotics (Fong, Nourbakhsh, & Dautenhahn, 2003). Furthermore, Scholtz (2002) and Kiesler and Goetz (2002) discussed the evaluation of assistive tools, and proposed the impact on caregivers (i.e., using robots for enhancing the work conditions of caregivers) and user satisfaction with received care from robots, as important benchmarks to consider in socially assistive robotics.

On the other hand, cognitive robotics has witnessed a big leap forward in intelligent frameworks and technologies. These recent frameworks and platforms include:

operating systems (e.g., BrainOS, ROS Quigley et al., 2009, and Urbi Baillie, 2004), perception (e.g., OpenCV and PCL library Rusu & Cousins, 2011), simulation (e.g., Webots and Gazebo Koenig & Howard, 2004), robot control interface (e.g., Orocos and Player Gerkey et al., 2001), and middleware (e.g., YARP Metta, Fitzpatrick, & Natale, 2006). These example frameworks contribute effectively to the current research of cognitive robotics in order to shape the future of human-robot interaction by endowing robots with advanced cognitive functions and better interaction services.

In this paper, we present a subjective analysis for some metrics, benchmarks, frameworks, and recent technologies in cognitive robotics, in addition to a summary of an interactive discussion session, with five expert researchers in the field, in order to draw up general and practical guidelines for bridging the gap between robotics research and applications. The rest of the paper is structured as following: Section 2 presents a detailed analysis of questionnaire data, Section 3 discusses the outcome of the study, Section 4 provides a summary of the panel discussion session, and last but not least, Section 5 concludes the paper.

2. Analysis and results

We planned to map out the emerging field of "intelligent social robots". For this purpose we compiled a questionnaire, which was distributed among the participants during the workshop "Towards Intelligent Social Robots – Current Advances in Cognitive Robotics" on the 3rd of November 2015 in Seoul, South Korea. In order to obtain a larger number of responses after the workshop, the URL for the online version of the same questionnaire was also passed on to other colleagues in the field. Overall, a total of 50 participants filled in the questionnaire. The responses are analyzed below. We must note that we do not claim that these are necessarily representative of the field as a whole, but merely indicative of the topics and trends within that field.

2.1. Countries of origin of the respondents

Although attendance at the larger robotics conferences usually does not depend on where a conference is held with respect to where participants are from (Griffiths, Voss, & Rohrbein, 2014), the majority of respondents for the questionnaire came from Europe (Czech Republic, Denmark, France, Germany, Italy, Romania, Spain, Switzerland, and the United Kingdom) with 60%. The second largest group are from Asian countries (China, Japan, Kazakhstan, South Korea, and Thailand) with 24%, followed by the Americas (Colombia and USA) with 12% and a small group of unknown origin (4%). These numbers are also displayed in Fig. 1. One needs to note that most respondents are from Europe. Consequently, a slight "Eurocentric" bias in the answers needs to be suspected.
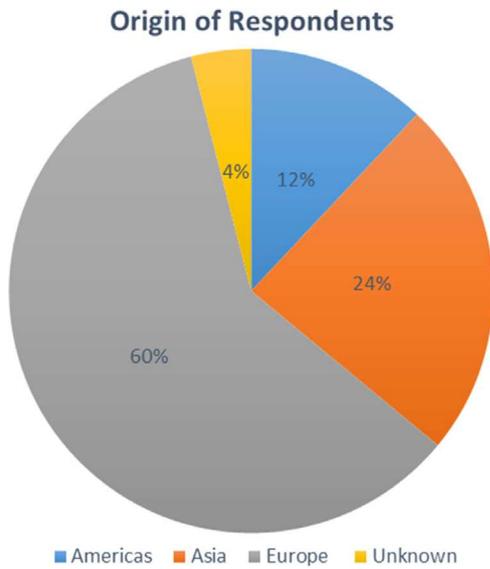
Fig. 1. The origin of respondents to the questionnaire according to continents.

Fig. 2. The disciplines to which the respondents report to belong to.

2.2. Respondents and their disciplines

"Intelligent Social Robotics" stands at the intersection of several disciplines. Thus, one very interesting question is which disciplines our participants identify with primarily. Here, we find some variation in granularity, which shows that one is dealing with an interdisciplinary community identified with fields such as computer science, cognitive science neuroscience, and psychology. However, most responses are on a finer level; whereas some respondents simply answered robotics, many responses were at the level of such sub-disciplines as cognitive robotics (Levesque & Lakemeyer, 2010), social robotics (Breazeal, Takanishi, & Kobayashi, 2008), or developmental robotics (Cangelosi, Schlesinger, & Smith, 2015). However, fields such as artificial intelligence or computer vision are also named repeatedly, and these do not necessarily stand directly "under" robotics in a taxonomy. A majority of participants identify with human-robot interaction (28%), which can certainly be seen as a sub-discipline of robotics in its own right.

The results are summarized in Fig. 2. We see the field of "intelligent social robotics" as allied mainly with cognitive robotics, which is the second largest distinct group (8%) in our survey. This was also suggested by the full title of the workshop. The combination of a variety of methods has become a common standard in cognitive robotics research (Rohlfing, Wrede, & Sagerer, 2014). However, we see this "sub-sub-discipline" of modern robotics as mainly focused on interaction and artificial sociality.

In the recent "survey of expert opinion" (Müller & Bostrom, 2014), one investigated question was about the research approach that would lead to what the authors call "high-level machine intelligence" (HLMI). Almost half of the respondents (47.9%) answered this question with "cognitive science". Only 2% of respondents in our survey saw their field as "cognition" in contrast to this.
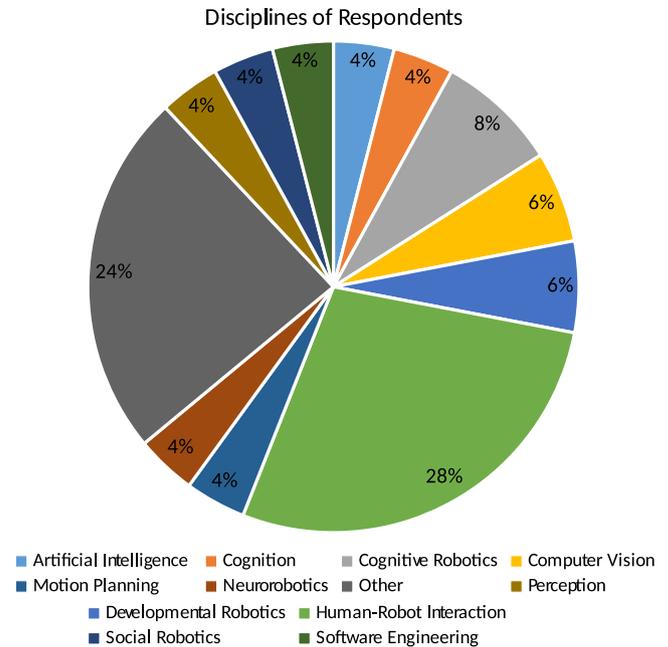
Gomila and Müller (2012) define 30 "measures of progress" with respect to the development of artificial cognitive systems. One of the areas in which the authors expect development is "social cognitive system", which also correspond to our area of interest under investigation.

For secondary affiliation with respect to disciplines the case is slightly less transparent. Some respondents gave more than one discipline keyword, but the majority (58%) decided to only give one discipline with which they affiliate. 16% of participants listed even more than two discipline keywords.

The non-primary disciplines again included human-robot interaction (named by 10% of respondents), control (4%) artificial intelligence (4%) and cognitive robotics (4%). Other disciplines were only named once. Most categories in both primary and non-primary fields of research named in this question would relate to the technical areas of robotics and computer science. However, there were also disciplines such as psychology or science and technology studies.

2.3. Paradigms for artificial cognitive systems

In line with Vernon (2014) and Frankish and Ramsey (2014), we offered the options of cognitivist approaches (also known as symbolic or knowledge based approaches) (Boden, 2014), connectionist approaches (also known as neural network modeling) (Sun, 2014), and dynamical systems (Beer, 2014). We also offered an "other" option. The results are displayed in Fig. 3.

The results show that the paradigms are almost balanced. All three possible paradigms used the entire spectrum with maximum values 7 being given and minimum values at 1. The cognitivist paradigm has a mean of
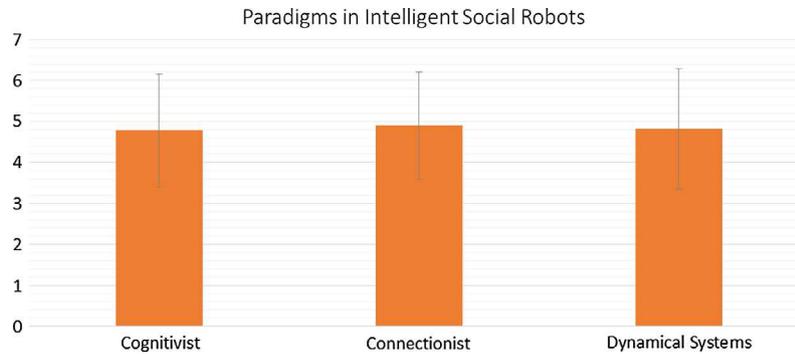
Fig. 3. Paradigms in intelligent social robots research.

$M = 4.78$ ($SD = 1.39$), the connectionist paradigm has a mean of $M = 4.9$ ($SD = 1.33$) and the dynamical systems paradigm has a mean of $M = 4.82$ ($SD = 1.49$). Although, the mean values are very close together the dynamical systems paradigm engendered slightly higher variability with respect to the responses. The median for all three paradigms was $Mdn = 5$, which again indicates that the respondents rated all three paradigms with very similar importance. However, with respect to the mode, the weakest measure of central tendency, the cognitivist paradigm has a mode of 6, while the dynamical systems paradigm has a mode of 5 and the connectionist paradigm has a mode of 4. The results show very little variance in the answers with a slightly more 'unified' opinion with regard to neural network modeling and more high values given for the cognitivist/symbolic paradigm. Overall, all three approaches seem to have their place within the field.

2.4. Contributions to future progress in intelligent social robotics

We were interested in what kinds of recent contributions to robotics were regarded as important to the scene. We hence offered the options ROS – the robot operating system (Quigley et al., 2009), the semantic web (Berners-Lee, Hendler, & Lassila, 2001), cloud robotics (Goldberg & Kehoe, 2013), and the multi-agent systems programming language (Bordini et al., 2006). These have all been named as significant new contributions in the literature. We also offered an "other" option. The results are displayed in Fig. 4.

ROS has a mean of $M = 4.7$ ($SD = 1.49$), the semantic web has a mean of $M = 4.3$ ($SD = 1.42$), cloud robotics has a mean of $M = 5.2$ ($SD = 1.32$), and the multi-agent systems programming language has a mean of $M = 4.24$ ($SD = 1.59$).

2.5. Benchmarking the social and cognitive skills of robots

As possible performance benchmarks, we offered the options of the Turing Test (Turing, 1950; Shieber, 1994; French, 2000; Harnad & Scherzer, 2008), Winograd schemas (Levesque, Davis, & Morgenstern, 2011), RoboCup

Soccer (Kitano, Asada, Noda, & Matsubara, 1998), and RoboCup@Home (Wisspeintner, Van Der Zant, Iocchi, & Schiffer, 2009). We also offered an "other" option. The results are displayed in Fig. 5.

The Turing test has a mean of $M = 3.92$ ($SD = 1.61$), the Winograd schema has a mean of $M = 3.98$ ($SD = 1.44$), RoboCup Soccer has a mean of $M = 3.36$ ($SD = 1.4$) and RoboCup@Home has a mean of $M = 4.24$ ($SD = 1.59$).

These data indicate that the respondents favoured RoboCup@Home over other benchmarking measures of intelligence. In line with this, among the "other" responses one respondent named RoboCup@Home as the single best measure of progress.

2.6. Frameworks for evaluation

Social and cognitive robotics have recently seen an expansion with respect to evaluation methods. Many of these are inspired by psychology, a field which has overall exerted quite a lot of control over the methodological framework of cognitive sciences in general in the past decades (Barsalou, 2010). In order of discriminating between different methodological frameworks we offered the choices of questionnaires, neuroscientific methods, behavioral methods, and system benchmarks. The results are displayed in Fig. 6.

Questionnaires have a mean of $M = 3.84$ ($SD = 1.58$), neuroscientific methods have a mean of $M = 4.48$ ($SD = 1.37$), behavioral methods have a mean of $M = 5.74$ ($SD = 0.9$) and system benchmarks have a mean of $M = 4.48$ ($SD = 1.37$). Clearly, behavioral methods were chosen as the most useful for the kind of research our respondents are interested in.

2.7. Metrics of human performance

In many tasks for cognitive robots in interaction with humans one needs to determine how well the human is doing at the joint task. For this purpose one needs metrics of performance. The questionnaire offered the choices of mental models, workload, and situation awareness. The results are displayed in Fig. 7.
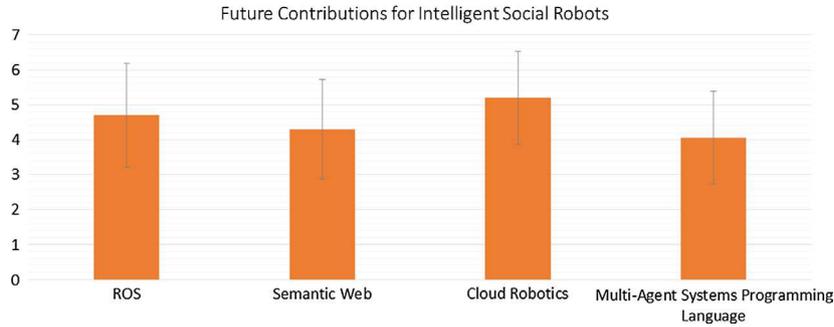
Fig. 4. Contributions to future progress in intelligent social robots research.
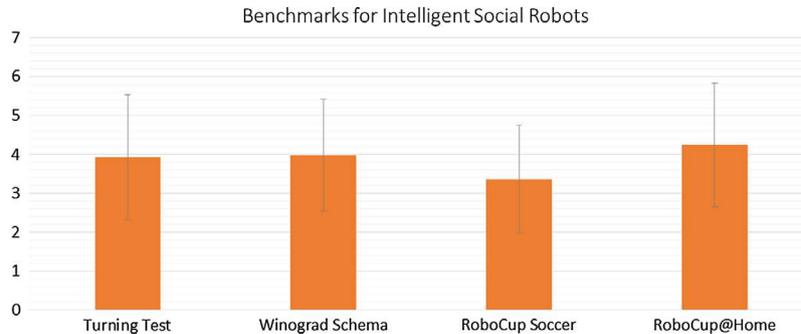
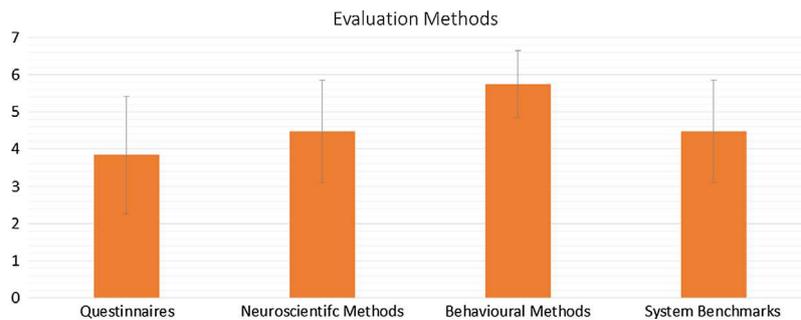

Fig. 5. Benchmarks in intelligent social robots research.

Fig. 6. Evaluation methods in intelligent social robots research.

Mental models have a mean of $M = 4.88$ $(SD = 1.22)$, workload has a mean of $M = 4.54$ $(SD = 1.42)$ and situation awareness has a mean of $M = 5.46$ $(SD = 1.42)$. The respondents showed a clear preference for situation awareness.

2.8. Metrics of robot performance

In almost all tasks for cognitive robots one needs to determine how well the robot is doing at a given task. For this purpose one needs metrics of performance. The questionnaire offered the choices of self-awareness, human-awareness, and autonomy. The results are displayed in Fig. 8.

Self-awareness has a mean of $M = 4.92$ $(SD = 1.34)$, human-awareness has a mean of $M = 5.42$ $(SD = 1.11)$ and autonomy has a mean of $M = 5.54$ $(SD = 1.31)$. Autonomy emerged as the favored response while self-wareness received the lowest average score.

2.9. Human-robot interaction metrics

As a means of measuring the effectiveness of human-robot interaction the questionnaire offered the options persuasiveness, engagement, and appropriateness of interaction. The results are displayed in Fig. 9.

Persuasiveness has a mean of $M = 4.72$ $(SD = 1.25)$, engagement has a mean of $M = 5.54$ $(SD = 1.11)$ and appropriateness of interaction has a mean of $M = 6.1$ $(SD = 0.86)$. The clear favorite of the respondents was appropriateness of interaction.

2.10. How do you evaluate the (current) scientific cooperation between the AI and HRI communities?

The mean answer to how well the artificial intelligence and the human-robot interaction communities interact is $M = 3.86$ $(SD = 1.21)$. Respondents did not use the full
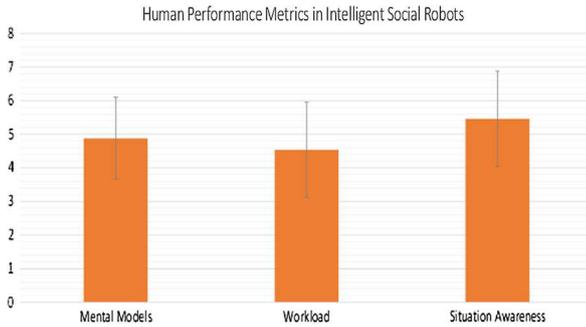
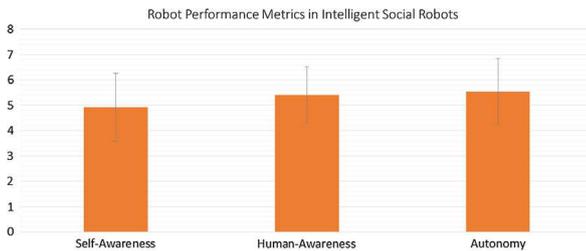Fig. 7. Human performance metrics in intelligent social robots research.

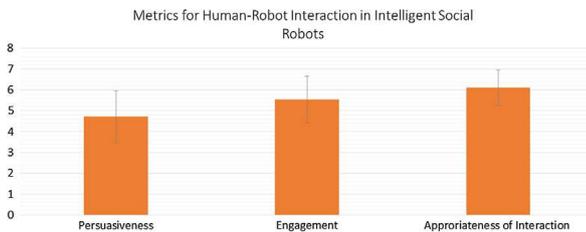Fig. 8. Robot performance metrics in intelligent social robots research.

Fig. 9. Metrics for human-robot interaction in intelligent social robots research.

scale with the maximum value given being 6 with 7 being the highest possible value and the minimum value being the lowest possible value 1. The median is $Mdn = 4$ and the mode is also 4. Thus, although there is room for improvement, the numbers indicate a rather positive picture with respect to this question.

What the questionnaire did not ask was whether this interaction is considered "useful" or "desired". However, artificial intelligence and robotics have been close since the beginning. However, cloud robotics (Waibel et al., 2011; Hunziker, Gajamohan, Waibel, & D'Andrea, 2013; Kehoe, Patil, Abbeel, & Goldberg, 2015) was also named repeatedly.

2.11. Major breakthroughs in intelligent social robots in the past five years

Participants were asked to name breakthroughs in "intelligent social robotics" of the past five years. We distinguished between breakthroughs in research and commercial developments. The answers to this question were open. The respondents showed repeatedly mentioned

deep learning (Coates et al., 2013; Schmidhuber, 2015; LeCun, Bengio, & Hinton, 2015) as one major breakthrough on the research level.

On the commercial level, the most named breakthrough is "Jibo"[2]. "Jibo" is a commercial product and the first social robot to be produced for the consumer market. However, the answers also included Pepper[3], NAO (Shamsuddin et al., 2011), and Baxter (Guizzo & Ackerman, 2012). However, the disembodied agent "Siri" and the speech recognition skills of personal assistant systems were also mentioned repeatedly as commercial breakthroughs.

3. Discussion

In the following section, we will discuss the results with respect to what we learned from the questionnaire. Several questions allowed an option for "other" where participants could answer free text. These answers are selectively used to illustrate some points below.

3.1. Building systems

We tried to gauge an answer as to how to build artificial social cognitive systems. We proposed several answers with respect to conventional cognitive architectures. No clear answer could be found with respect to whether neural networks, knowledge-based systems or dynamical systems are preferred among the community.

Interestingly, however, the answers which were obtained with respect to "other", often pointed toward hybrid approaches or "gray box" (Zorzetto, Maciel Filho, & Wolf-Maciel, 2000) approaches, which roughly mean hybrid. Further, some respondents pointed toward the task dependent nature of the solution. Thus, the responses we obtained suggest that the community at the moment seems to favor a pragmatic approach, over favoring a certain paradigm. Instead of a one-size fits all model of cognition, task-dependent designs or hybrid systems are named as the favored approaches in the future.

Participants also see the exchange between artificial intelligence and cognitive robotics as alive and well. Thus, moving forward we can assume that the current developments in this field will also influence developments in "intelligent social robots". The main goal is to build autonomous systems, which posses the right capabilities for interaction.

3.2. Evaluating systems

When it comes to evaluating systems, the pool of respondent again favored an overall pragmatic approach. The answers tend to stress the need for practical tasks and behavioral measures.

The participants saw RoboCup@home as a valuable means of benchmarking performance. They preferred behavioral measures over questionnaires, neuroscientific methods and even system benchmarks. Also, human performance should best be measured in terms of situation awareness. Robots in contrast should be judged by their degree of autonomy. The overall goal in interaction is to see whether the degree of "appropriateness of interaction" is satisfied.

3.3. Deploying systems

Among the benchmarking responses one finds suggestions that "long-term interaction" in situ may be the key to social cognition for robots. Thus, if one can build learning systems these can ultimately become social over time. This idea also agrees well with the relative popularity of the RoboCup@home competition as a benchmark.

For the group of respondent the interaction with robots in naturalistic scenarios appears to be the natural way forward with respect to "intelligent social robots". This also was clear from answers in the "other" category of to benchmarking question, which included "daily interaction with end users" and robots deployed in long-term setting and in families. It was pointed out repeatedly that the overall behavior of a robot would be more important than its performance at a given isolated task (e.g. conversation). The community seems to be interested in testing robots in naturalistic scenarios more than on an abstract level. This needs to be considered moving forward.

3.4. Building a community

The question of whether the "intelligent social robots" community is interlinked with a variety of other communities was also addressed. The respondents in our questionnaire affiliate with various disciplines and sub-disciplines of robotics, but some are also affiliated with psychology and other social sciences. Generally, one can see the community has a strong association with human-robot interaction, but the community is also especially interested in the cognitive aspects of robotics.

4. Panel discussion summary

In this section, we provide a summary of the panel discussion session with five renowned researchers in cognitive robotics, whose experiences vary from industry to academia:

- **Dr. Amit Kumar Pandey - Aldebaran Robotics - France** is a chief scientist at Aldebaran Robotics in Paris, France, serving as the scientific coordinator (R&D) of the collaborative projects. He holds a bachelor's degree from Jaypee Institute of Information Technology and a master's degree from the Indian Institute of Technology (IIT) in India, and he received a doctorate from LAAS-CNRS in France. Dr. Pandey's research interests include socially intelligent robots and societal applications, human-robot interaction, and artificial intelligence-based applications.

- **Prof. Jun Tani - Korea Advanced Institute of Science and Technology (KAIST) - South Korea** is a full-professor at the Department of Electrical Engineering and the director of the Cognitive Neuro-Robotics Laboratory at the Korea Advanced Institute of Science and Technology (KAIST) in Daejeon, South Korea. He received his bachelor's degree from Waseda University in Japan, a master's degree from the University of Michigan in USA, and a doctorate from Sophia University in Japan. He started his career as a senior researcher at the Sony Computer Science Laboratory in Japan, afterwards he was appointed as a team leader in the Behavior and Dynamic Cognition Laboratory at RIKEN Brain Science Institute in Japan. Prof. Tani's research interests focus on understanding the underlying mechanisms of embodied cognition and mind through conducting an interdisciplinary research that combines robotics, brain science, non-linear dynamics, and psychology.

- **Prof. Sonya Kwak - Ewha Womans University - South Korea** is an industrial design professor at the Ewha Womans University in Seoul, South Korea. She completed her undergraduate, master's, and Ph.D. degrees at the department of industrial design in the Korea Advanced Institute of Science and Technology (KAIST) in South Korea. Prof. Kwak's research interests include human-robot interaction based on social psychology, in addition to designing human-friendly robots (e.g., Mung, Haemi, and Hangul-bot robots).

- **Prof. Yukie Nagai - Osaka University - Japan** is a specially appointed associate professor at the Emergent Robotics Laboratory of the Graduate School of Engineering at Osaka University in Osaka, Japan. She received her bachelor's and master's degrees from Aoyama Gakuin University and her doctoral degree from Osaka University in Japan. Afterwards, she worked as a postdoctoral researcher at Bielefeld University in Germany before returning to Osaka University. Prof. Nagai does research on cognitive developmental robotics and her research interests include cognitive development based on predictive learning, infant-directed action, joint attention, and human-robot interaction.

- **Prof. Lorenzo Natale - Italian Institute of Technology (IIT) - Italy** is a tenure track researcher at the iCub Facility Department of the Istituto Italiano di Tecnologia (IIT) in Genoa, Italy. He received his undergraduate, master's, and doctoral degrees from the University of Genoa in Italy. Thereafter, he was a postdoctoral research associate at the MIT Computer Science and Artificial Intelligence Laboratory in USA before returning to IIT Institute as the leader of the Humanoid Sensing and Perception research group. Prof. Natale's research interests range from sensorimotor learning and vision, to software engineering and system integration in robotics.

In this rich discussion session, the invited panelists exchanged opinions actively with the workshop attendees about different robotics-related topics and challenges. In the following points, we summarize some of the addressed questions to the panelists and highlight their corresponding points of view:

1. **In the recent European Commission review (European Robotics Research: Achievements and Challenges) issued in 2012, the commission stated the need to fill the gap separating research and innovation**. We asked our panelists about the potential causes and their recommendations to fill this gap:

   - **Pandey**: A possible way to fill the gap separating research and innovation is to encourage partnerships between research and industry. Researchers should bring the results of their studies forward to industrial partners, who would evaluate, in turn, any potential outcome for commercial and innovative products. Although some research projects produce innovative ideas and technologies, there might be practical problems in embedding and employing such technologies in commercial robots and applications. This could be related to the cost of the technology, the computational effort it requires, etc. The considered Technological Readiness Level (TRL) for H2020 projects will be, probably, useful to fill the gap separating research and innovation. Funding research projects that meet, at least, the minimum level of TRL, would certainly enhance the chances of having strong results that could be used for building innovative commercial applications.

   - **Tani**: Researchers are not supposed to work on direct and immediate applications for the industry. Technology comes out bottom-up, and not because the governments are expecting a specific outcome or putting constraints on researchers for creating it. For example, thinking about the recent unexpected breakthrough of deep learning, we had the suspicion that Recurrent Neural Networks (RNNs) were about to die. However, in the last two years, RNNs are really running deep learning. Consequently, we should consider the gradual development of technology and innovation.

   - **Kwak**: Researchers should understand the needs of users so as to propose ideas on how to create technology and build products that satisfy these needs. There is even too much technology that we do not know where and how to use. Researchers should try to understand how to deliver the available technology to objects that people use in daily life.

   - **Nagai**: A possible way to fill the gap separating research and innovation is to foster the discussion among scientists working on basic and applied sciences. For example, researchers working on cognitive robotics and researchers working on social robotics should make a real effort to understand how to combine and integrate the results of cognitive science research studies in order to build social robots-based applications that could satisfy the different needs of users.

   - **Natale**: The new trend taken by the EU Commission in structuring innovation more than basic science might have been caused by the apparent inability of basic science researchers to understand how to transform the findings of their studies into real commercial products that people can use. For bridging the gap between research and innovation, I think that researchers working on basic science should foster the discussion with people working in the industry in order to understand the problems they face and try to come up with solutions to their practical problems.

   - **Audience**: Robotics research should change its perspective; rather than focusing on building simple robot companions for example, it should work on task oriented platforms and robots that can engage effectively in short term interactions. A positive example of such a direction is represented by a robot like Pepper[4], which could be considered as a real innovative product that can be used in people's social life in a wide range of circumstances.

2. **A human cognitive system could be defined in terms of a group of fundamental cognitive processes, like: knowledge acquisition, storing, and processing, in addition to learning from experience, which allow humans to generate adapted behaviors**. We asked our panelists how far we are from endowing robots with an artificial cognitive system of a similar functionality:

   - **Pandey**: For building robots with cognitive capabilities, we need to bring them through a developmental process similar to that of children. We should give them curiosity to explore the world and the capability to learn from sensorimotor inputs. However, this process is so sophisticated and needs a long time. We can either give robots time to develop (i.e., through a bottom-up development), or embed some developmental blocks through which the robot can start acquiring more knowledge to learn. Aldebaran Robotics is currently investigating both approaches and any possible trade-off.

   - **Tani**: How close are robots to humans? For example, IBM Watson - the question answering computer system - is very good at answering general questions similarly to humans. But, Watson can never answer the questions that require reasoning about sensory information (e.g., the smell of objects). At some limited level of linguistic interaction, Watson could look close to a real human. Coming back to the question, it depends on how intelligence is measured; a simple

---

4 https://www.ald.softbankrobotics.com/en/cool-robots/pepper.

program that can perform a simple task could be considered intelligent. However, speaking about human intelligence - with all the associated and sophisticated cognitive processes - in a general scope, I guess it would be very difficult to endow robots with a cognitive system of a similar functionality.

- **Kwak**: Do we really need a human-like technology? Where are we going to use it? If we make a perfect human-like robot, how much will people pay for it? We will need to fit the profit with the cost of materials so as to make the product at a good price, and a good example on that is Pepper robot. I hope that our community thinks about the cost and function of technology so as to offer affordable products that normal people can get.

- **Nagai**: Many researchers today are working on motivation models, not only on learning models. Through these models, robots can increase their knowledge and autonomously select what to learn depending on what they can do, because if the targets are so difficult, they might not be able to gain any learning experience similarly to humans. However, if the targets are close to their existing knowledge, they can gain new experiences. Therefore, we can design motivation/curiosity models so as to make robots autonomously select what to do and what to learn. Consequently, integrating learning and motivation models could make robots undergo continuous cognitive development as humans.

- **Natale**: I believe that we are very far from endowing robots with a similar artificial cognitive system. Robots execute some defined tasks through programs written by humans. Robotics research needs to find the way to enable robots to do new tasks, and come up with new behaviors/actions autonomously. Robots can incrementally improve what they were programed to do, but still cannot totally behave independently.

3. **Theory-of-mind could be defined as the ability to attribute mental states (i.e., belief, desire, and intention) to other people/agents so as to understand the others' beliefs, desires, and intentions that could be different from one's own**. We asked our panelists about the possibility of endowing a robot with a theory-of-mind based cognitive model so as to predict another human/robot's actions:

- **Pandey**: Theory-of-mind between robots and humans is being covered by research that shows some good results. However, applying this theory on two robots would be, indeed, very interesting. In terms of the implementation and development side, if the implementation is generic enough to treat an agent either as a robot or a human in the same way, then we can have this kind of cognitive capability easily adopted between two robots.

- **Tani**: I cooperate with Prof. Murata from Waseda University on making two robots that learn from each other using predictive learning. When two robots become able to predict each other's behavior, they start to collaborate. Murata's model is responsible for detecting the possible level of prediction. In case the robot realizes that it cannot predict another robot's behavior, it starts doing its own things. On the contrary, in case the robot realizes that it can predict another robot, they start collaborating. Therefore, it is important to be able to understand when prediction is possible, and when it is not, so that one makes the robot able to collaborate with another one.

- **Nagai**: What is the definition of theory-of-mind? My definition for theory of mind is the ability to use our internal models to understand the environment or the world. Based on this definition, humans could apply theory of mind to understand robots, and robots could use their internal cognitive models to understand humans. An interesting example are autistic people, who also have the ability to apply theory-of-mind to understand others, even though several researchers argue that they cannot read the internal states of other persons. The problem is that their internal models are different from those of typical persons. That is why they cannot communicate with typical persons. However, they can communicate very well with people with ASD (Autism Spectrum Disorder), because they share similar internal models. Consequently, if robots can acquire similar internal models to those of humans, they would be able to understand what humans are thinking about, and similarly humans would be able to understand robots, like understanding why a robot is not working.

- **Natale**: I think that applying a theory-of-mind between robots is feasible. You can always access the program of a robot, so that you can predict some its behaviors, knowing how it is programmed. An expected problem that would be present in this context is the perception problem, like detecting where the robot is looking through a perspective-taking; as you are in a different perspective than that of the robot. Afterwards, in terms of predicting how the robot would behave in situations, you do not need a learning model for that, you can anticipate some behaviors if you can access the program, so that you can expect how the robot would behave in response to stimuli for example. However, in order to make the robot more autonomous, it should generate some other random behaviors that would need complex cognitive models to be predicted.

5. Conclusion

This paper discusses the outcome of our first annual workshop: *Towards Intelligent Social Robots - Current Advances in Cognitive Robotics*, which was held in

conjunction with the 15th IEEE-RAS Humanoids Conference - Seoul - South Korea - 2015. In this technical report, we illustrate the findings of a subjective analysis for some metrics, benchmarks, and frameworks in cognitive robotics. Additionally, we provide a summary of an interactive discussion session, with five expert researchers, about different robotics-related topics, which could help in enriching the vision of young researchers for the awaiting future challenges. In the future versions of our workshop, we will continue analyzing and reporting the different challenges in cognitive robotics with cutting-edge discussions with expert researchers in the field.

References

Baillie, J. (2004). URBI: Towards a universal robotic body interface. In *Proceedings of the 4th IEEE/RAS international conference on humanoids robots (Humanoids), CA, USA*.

Barsalou, L. W. (2010). Introduction to 30th anniversary perspectives on cognitive science: Past, present, and future. *Topics in Cognitive Science, 2*(3), 322–327.

Beer, R. D. (2014). Dynamical systems and embedded cognition. In K. Frankish & W. M. Ramsey (Eds.), *The Cambridge handbook of artificial intelligence* (pp. 128–148). Cambridge, UK: Cambridge University Press.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American, 284*(5), 28–37.

Boden, M. A. (2014). GOFAI. In K. Frankish & W. M. Ramsey (Eds.), *The Cambridge handbook of artificial intelligence* (pp. 89–107). Cambridge, UK: Cambridge University Press.

Bordini, R. H., Braubach, L., Dastani, M., Seghrouchni, A. E. F., Gomez-Sanz, J. J., Leite, J., ... Ricci, A. (2006). A survey of programming languages and platforms for multi-agent systems. *Informatica, 30*(1).

Breazeal, C., Takanishi, A., & Kobayashi, T. (2008). Social robots that interact with people. In *Springer handbook of robotics* (pp. 1349–1369). Heidelberg: Springer.

Cangelosi, A., Schlesinger, M., & Smith, L. B. (2015). *Developmental robotics: From babies to robots*. Cambridge, MA: MIT Press.

Coates, A., Huval, B., Wang, T., Wu, D., Catanzaro, B., & Andrew, N. (2013). Deep learning with COTS HPC systems. In *Proceedings of the 30th international conference on machine learning* (pp. 1337–1345).

Feil-Seifer, D., Skinner, K., & Mataric, M. (2007). Benchmarks for evaluating socially assistive robotics. *Interaction Studies: Psychological Benchmarks of Human-Robot Interaction, 8*(3), 423–429.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems, 42*(3–4), 143–166.

Frankish, K., & Ramsey, W. M. (Eds.). (2014). *The Cambridge handbook of artificial intelligence*. Cambridge: Cambridge University Press.

French, R. M. (2000). The Turing test: The first 50 years. *Trends in Cognitive Sciences, 4*(3), 115–122.

Gerkey, B., Vaughan, R., Stoy, K., Howard, A., Sukhatme, G., & Mataric, M. (2001). Most valuable player: A robot device server for distributed control. In *Proceedings of the 14th IEEE/RSJ international conference on intelligent robots and systems (IROS), Maui, USA*.

Goldberg, K., & Kehoe, B. (2013). Cloud robotics and automation: A survey of related work. Tech. Rep. UCB/EECS-2013-5, EECS Department, University of California, Berkeley, Berkeley, California.

Gomila, A., & Müller, V. C. (2012). Challenges for artificial cognitive systems. *Journal of Cognitive Science, 13*, 453–470.

Griffiths, S., Voss, L., & Rohrbein, F. (2014). Industry-academia collaborations in robotics: Comparing Asia, Europe and North-America. In *Proceedings of the IEEE international conference on robotics and automation (ICRA)* (pp. 748–753).

Guizzo, E., & Ackerman, E. (2012). How rethink robotics built its new Baxter robot worker. In *IEEE spectrum*.

Harnad, S., & Scherzer, P. (2008). First, scale up to the robotic Turing test, then worry about feeling. *Artificial Intelligence in Medicine, 44*(2), 83–89.

Hoffman, G. (2013). Evaluating fluency in human-robot collaboration. In *Proceedings of the international conference on robotics: Science and systems (RSS), Berlin, Germany*.

Hunziker, D., Gajamohan, M., Waibel, M., & D'Andrea, R. (2013). Rapyuta: The RoboEarth cloud engine. In *Proceedings of the IEEE international conference on robotics and automation* (pp. 438–444) http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6630612, doi:http://dx.doi.org/10.1109/ICRA.2013.6630612.

Kahn, P., Ishiguro, H., Friedman, B., Kanda, T., Freier, N., Severson, R., & Miller, J. (2007). What is a human? Toward psychological benchmarks in the field of human-robot interaction. *Interaction Studies: Psychological Benchmarks of Human-Robot Interaction, 8*(3), 363–390.

Kehoe, B., Patil, S., Abbeel, P., & Goldberg, K. (2015). A survey of research on cloud robotics and automation. *IEEE Transactions on Automation Science and Engineering, 12*(2), 398–409.

Kiesler, S., & Goetz, J. (2002). Mental models and cooperation with robotic assistants. In *Proceedings of the 20th ACM international conference on human factors in computing systems (CHI), Minnesota, USA*.

Kitano, H., Asada, M., Noda, I., & Matsubara, H. (1998). RoboCup: Robot world cup. *Robotics & Automation Magazine, IEEE, 5*(3), 30–36.

Koenig, N., & Howard, A. (2004). Design and use paradigms for Gazebo: An open-source multi-robot simulator. In *Proceedings of the 17th IEEE/RSJ international conference on intelligent robots and systems (IROS), Sendai, Japan*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521* (7553), 436–444. http://dx.doi.org/10.1038/nature14539http://dx.doi.org/10.1038/nature14539/n10.1038/nature14539. Available from: arXiv:arXiv:1312.6184v5.

Levesque, H. J., & Lakemeyer, G. (2010). Cognitive robotics. In G. Lakemeyer, H. J. Levesque, F. Pirri (Eds.), *Cognitive robotics, no. 10081 in Dagstuhl seminar proceedings, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany, Dagstuhl, Germany*. <http://drops.dagstuhl.de/opus/volltexte/2010/2633>.

Levesque, H. J., Davis, E., & Morgenstern, L. (2011). The Winograd schema challenge. In *Proceedings of the AAAI spring symposium: Logical formalizations of commonsense reasoning*.

Metta, G., Fitzpatrick, P., & Natale, L. (2006). YARP: Yet another robot platform. *Advanced Robotic Systems, 3*(1), 43–48.

Müller, V. C., & Bostrom, N. (2014). *Future progress in artificial intelligence: A survey of expert opinion*. Berlin: Springerhttp://www.sophia.de/pdf/2014_PT-AI_polls.pdf.

Olsen, D., & Goodrich, M. (2003). Metrics for evaluating human-robot interactions. In *Proceedings of the 4th international workshop on performance metrics for intelligent systems (PERMIS), Maryland, USA*.

Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., ... Ng, A. (2009). ROS: An open-source robot operating system. In *Proceedings of the international workshop on open source software - In conjunction with the IEEE international conference on robotics and automation (ICRA), Kobe, Japan*.

Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., ... Ng, A. Y. (2009). Ros: An open-source robot operating system. *ICRA workshop on open source software* (Vol. 3).

Rentschler, A., Cooper, R., Blasch, B., & Boninger, M. (2003). Intelligent walkers for the elderly: Performance and safety testing of VA-PAMAID robotic walker. *Rehabilitation Research and Development, 40*(5), 423–431.

Rohlfing, K. J., Wrede, B., & Sagerer, G. (2014). Dialog: Trained on everything. *AMD Newsletter, 11*(2), 4–5.

Rusu, R., & Cousins, S. (2011). 3D is here: Point cloud library (PCL). In *Proceedings of the IEEE international conference on robotics and automation (ICRA), Shanghai, China.*

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85–117. http://dx.doi.org/10.1016/j.neunet.2014.09.003http://arxiv.org/abs/1404.7828. Available from: arXiv:arXiv:1404.7828v1.

Scholtz, J. (2002). Evaluation methods for human-system performance of intelligent systems. In *Proceedings of the 3rd international workshop on performance metrics for intelligent systems (PERMIS), Maryland, USA.*

Shamsuddin, S., Ismail, L. I., Yussof, H., Zahari, N. I., Bahari, S., Hashim, H., & Jaffar, A. (2011). Humanoid robot NAO: Review of control and motion exploration. In *Proceedings of the IEEE international conference on control system, computing and engineering (ICCSCE)* (pp. 511–516), doi:http://dx.doi.org/10.1109/ICCSCE.2011.6190579.

Shieber, S. M. (1994). Lessons from a restricted Turing test. *Communications of the Association for Computing Machinery*(6), 70–78.

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006). Common metrics for human-robot interaction. In *Proceedings of the 1st ACM/IEEE international conference on human-robot interaction (HRI), Utah, USA.*

Sun, R. (2014). Connectionism and neural networks. In K. Frankish & W. M. Ramsey (Eds.), *The Cambridge handbook of artificial intelligence* (pp. 108–127). Cambridge, UK: Cambridge University Press.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*, 433–460.

Vernon, D. (2014). *Artificial cognitive systems: A primer*. Cambridge, USA: MIT Press.

Waibel, M., Beetz, M., Civera, J., D'Andrea, R., Elfring, J., Gálvez-López, D., Häussermann, K., Janssen, R., Montiel, J., Perzylo, A., Schießle, B., Tenorth, M., Zweigle, O., & van de Molengraft, R. (2011). RoboEarth. *IEEE Robotics & Automation Magazine, 18*(2), 69–82. http://dx.doi.org/10.1109/MRA.2011.941632.

Wisspeintner, T., Van Der Zant, T., Iocchi, L., & Schiffer, S. (2009). Robocup@Home: Scientific competition and benchmarking for domestic service robots. *Interaction Studies, 10*(3), 392–426.

Zorzetto, L. M., Maciel Filho, R., & Wolf-Maciel, M. (2000). Processing modelling development through artificial neural networks and hybrid models. *Computers & Chemical Engineering, 24*(2), 1355–1360.