Berlin Summit on Robotics

Conference Report 2011

December 2, 2011

Foreword

The Berlin Summit on Robotics took place in Berlin from July 20 until July 24, 2011. It brought together leading researchers in robotics and related areas to discuss and tackle strategic challenges in robotics research. While conferences and other professional meetings are hectic and dominated by conversations about technical details, the Berlin Summit aimed to create an atmosphere in which big ideas can spring to life and grow into strategic initiatives for the robotics community.

To make a free exchange of ideas as easy as possible, the program of the Berlin Summit did not adhere to a fixed schedule. Individuals or small groups of individuals prepared some topics for discussion and distributed brief "discussion primers" prior to the meeting. For each of these topics, this report contains a chapter. Each chapter is meant to capture the content of the discussion primers, the discussion at the workshop, and some conclusion or recommendations. We hope you will find the following chapters as stimulating as the participants found the discussions in Berlin.

We are greatly indebted to the Alexander von Humboldt Foundation for funding this event.

Participants

- **Tamim Asfour** | KIT, Germany
- **Oliver Brock** | TU Berlin, Germany
- **Wolfram Burgard** | Universität Freiburg, Germany
- **Kostas Daniliidis** | U Penn, USA
- **Andrew Davison** | Imperical College, UK
- **Greg Dudek** | McGill University, Canada
- **Oussama Khatib** | Stanford University, USA
- **Vijay Kumar** | U Penn, USA
- **Steve LaValle** | UIUC, USA
- **Cécile Huet** | European Commission, Luxembourg
- **Tomas Lozano-Perez** | MIT, USA
- **Matt Mason** | CMU, USA
- **Paul Newman** | Oxford, UK
- **Nick Roy** | MIT, USA
- **Stefan Schaal** | USC, USA
- **Roland Siegwart** | ETH, Switzerland
- **Gaurav Sukhatme** | USC, USA
- **Marc Toussaint** | FU Berlin, Germany

On the Conference

“Just wanted to say thanks again for organising this fantastic meeting!”

[Andrew Davison | Imperial College London]

“Let me thank you again for organizing the very rewarding meeting we just had. [. . .] Again, that’s for a pleasant, well-organized and fruitful event. It was a real triumph on your part.”

[Gregory Dudek | McGill University]

“Congratulations on an amazing meeting. I cannot think of a more stimulating meeting. I can only speculate that it will have some considerable impact in shaping the field. Also, it was fun. Thanks!”

[Matt Mason | Carnegie Mellon University]

“This is the second-best meeting I have been to. The best meeting took place in a castle in France with a wine cellar and a swimming pool; and there were three butlers and a cook pamper us.”

[Stefan Schaal | University of Southern California]

“[. . .] to say how much I enjoyed that week’s Berlin Summit. You did [. . .] our community a great service.”

[Paul Newman | University of Oxford]

“Three most refreshing days of honest introspection, fresh ideas, and food for thought for the whole year! The reading is an example par excellence of what is a big picture and a vision for a field.”

[Kostas Daniilidis | University of Pennsylvania]

Table of Contents

1 Is Robotics in Need of a Paradigm Shift?

Oliver Brock | *TU Berlin*

Paradigm shifts can lead to explosive progress in the Natural Sciences. A paradigm shift is a disruptive change in the basic assumptions held by the majority of scientists working in a field. Here, we examine the possibility that robotics must undergo such a paradigm shift to make profound progress towards truly autonomous or even intelligent robotic systems with human-like capabilities.

1.1 What is a paradigm shift? The story of Phlogiston

In 1667 Johann Joachim Becher postulated the existence of phlogiston, a massless, tasteless, odorless, and colorless substance, to provide an explanation for the phenomenon of combustion. Combustible materials, according to phlogiston theory, contain phlogiston which is released during combustion. Today, of course, we know that combustion is oxidation and it appears strange that the phlogiston theory would ever have been accepted. Nevertheless, it was the dominant scientific paradigm to explain combustion until the last third of the eighteenth century. And there were some merits to it. For example, it could explain why a candle extinguishes when placed in a confined space: according to the theory, the candle sets free phlogiston and once the air is saturated combustion stops. It could also explain why combustible materials would only burn for a certain time: once their phlogiston supply was depleted combustion stopped.

The phlogiston theory was the accepted explanation for combustion for about a century. The discovery that metals gain mass during combustion would ultimately lead to its demise. However, it took several decades until the majority of scientists gave up phlogiston and accepted the theory of oxidation.

This example is typical; there are many more examples: the switch from the geocentric to the heliocentric world model, or the numerous theories about gravity from Aristotle to Descartes, from Newton to Einstein. For many of the past paradigm shifts, there is evidence that well-known and respected scientists deliberately ignored scientific evidence and stuck to the “old” paradigm. Thomas Kuhn, who coined the term paradigm shift [4], pointed out in their defense that a paradigm shift is accompanied by fundamental changes in emphasis and perception. It is thus credible that the scientific community can remain stuck in an old paradigm until they are ready to change perspective and to embrace the new paradigm.

History of science thus tells us that we can get stuck in old paradigms that hamper progress and make us look stupid in retrospect. The least we should learn from this lesson is that, every once in a while, we should step back and try to see if we are stuck in a paradigm whose time has come.

1.2 Are there signs that we need a paradigm shift in robotics?

What are the signs of an imminent paradigm shift? In the natural sciences this question might be easier to answer and the example of phlogiston already gives us an idea. But before we can attempt to answer this question for robotics, we have to come to an agreement what we mean by robotics. For the purpose of this discussion, I would like to assume that robotics is the field that attempts to build embodied artificial intelligence—machines that can decide, act, learn, and transfer acquired knowledge between domains autonomously. These machines differ from today's robots in that they are to some degree universal and not limited to perform a narrowly defined task.

It is the very nature of a paradigm shift that it is difficult to see from within the existing paradigm. And we do not hope to offer a decisive answer here. All we might be able to do is to point to anecdotal signs. The reader then will have to decide what conclusions to draw from the evidence.

Thomas Kuhn wrote that "paradigm-testing occurs only after persistent failure to solve a noteworthy puzzle has given rise to crisis" [4]. So is robotics in a crisis? Is robotics experiencing a persistent failure to solve an important problem? From the very beginnings of robotics, it has been the goal to build robots that perform general tasks in general environments. An SRI Technical Memo from 1966 states as the objective of the Shakey project the development of "concepts and techniques in artificial intelligence enabling an automaton to function independently in realistic environments" [7]. To this day, this objective seems to be out of reach. Still in 2006 a program call of the National Science Foundation stated that the scope of targeted research "encompasses computational understanding and modeling of [...] capabilities that demonstrate intelligence and adaptability in unstructured and uncertain environments." It would seem that we have pursued the same goal for over forty years.

One could now begin a long debate. Of course, there has been substantial progress in robotics over the last forty years. We have Stanley and the Mars rover, for example. We have developed numerous foundational theories and algorithms. Planning, control, machine learning, computer vision to name but a few – they all have advanced tremendously. But the key question remains: are these advances making progress towards robots with human-like capabilities and task generality? Or are we building more and more advanced machines with super-human capabilities in a specific niche?

The European Commission in 2009 answered this question as follows: "Engineering systems with the capability to sense and understand an unstructured environment is a challenge which goes beyond today's systems engineering paradigm." [8] One might call this a crisis. And if there is even the slightest chance that this assessment is correct, we should question the existing paradigm and attempt to discover its successor.

1.3 What is the problem?

The expression "pixels to torques" might be the simplest way of describing the high-level agenda of research at the intersection of AI and robotics: How can we create an agent capable of taking all of its sensor input and turning into meaningful and purposeful motion to affect the world and perform active sensing? How can such an agent explore the world, learn from its mistakes, or apply past experiences to novel situations?

Clearly, our role model is the human. While roboticists would already be deeply impressed by an agent exhibiting the autonomy of a hedgehog, a look at the ultimate target might give us some ideas about the characteristics of the problem. Assuming that nature is parsimonious, the complexity of sensing and actuation might give us some idea of the "pixels to torques"-problem for humans. Of course, the human evolved in response to various survival pressures our robots will never be subjected to, but let's just see where this leads us.

Each human eye has approximately 120 million rods and 6 million cones, for a total of 232 million light-sensitive nerve ends. Forty million nerve endings are dedicated to the sense of smell, 3.5 million to the sense of touch. Each human ear has 15,000 to 20,000 auditory nerve receptors and the tough possesses 10,000 taste buds. So much for the input. The output is achieved by actuating between 650 and 850 muscles, depending on how you count. So overall, the human agent has 300 million inputs and 800 outputs (ignoring the skin's 200,000 sweat glands).

In robotics, there is a clear trend towards more sensors with higher resolutions and more degrees of freedom of the mechanism. Nevertheless, robotic systems are still far away from the human numbers. Are the mechanical and sensing systems of today's robots capable of performing human tasks at human-level performance? One might argue that a human is capable to perform almost any task teleoperating today's robots, even when limited in sensing to the robot's sensors. However, the question remains to be answered if a more advanced sensor suite that those found on today's robots is needed to acquire the skills necessary for this teleoperation.

To understand the problem solved by human agents it is not sufficient to look at the numbers at the interface between the agent and its environment. It is in fact the interactions across the interface that seem most relevant. There is ample evidence in psychology and neuroscience to support that this boundary (or interface) is not as strict as we might assume. Cognitive processes in humans are distributed across this boundary; a well-established example of this is using the environment as working memory rather than storing a detailed representation in the brain.

For our considerations this may mean that the interface is just the narrowest part of information flow within the agent, the world, and in between. The true complexity may lie in the interactions themselves. These interactions concern many objects with many properties to achieve many different objectives. Knowledge of the state of the world is uncertain and the ability to perform the desired action is also imperfect. In addition, the physics of multi-contact between objects is highly complex: small differences in initial conditions may have completely different outcomes.

All of this to arrive at a well-known conclusion: the problems we are trying to solve are extremely high-dimensional, inherently complex, and riddled with uncertainty. These problems might be graver than the robotics community has assumed or been willing to admit. Finding a solution might therefore require a paradigm shift. Are there any candidates for a winning paradigm?

1.4 What are candidates for paradigms in robotics?

1.4.1 System Engineering

The program managers at the European Commission refer to a "current systems engineering paradigm" and already express the opinion that it will not suffice to create autonomous and task-general robots [8]. Clearly, it is a gross oversimplification to speak of a single current paradigm. Nevertheless, there is value in trying to capture the dominant characteristics of today's research in robotics. This is what I will attempt to do in this section. It will be easy, of course, to find ongoing research activities that do not comply with these characteristics. But that is not the point.

The politics of the funding landscape today make it necessary for researchers to demonstrate the immediate applicability of their results to real-world problems. This, in my opinion, creates a strong bias for the type of questions that are being asked. Furthermore, this bias has a deep interdependency with the current paradigm.

To motivate research in the current paradigm, one has to have an application in mind. The application implies a decomposition of the targeted problem into sub-problems. The sub-problems are solved, possibly by different systems (vision system, mobility, etc.) and then integrated into a bigger system. Hence, we can speak of a systems engineering paradigm. This paradigm finds its apex in the App Store-idea: a software system warehouse from which everybody can obtain sub-systems and compose them into application-solving systems. There are obvious successes that justify pursuing the current paradigm. And there also seem to be signs of stagnations in other areas, justifying this document.

1.4.2 Computation

Available computational power continues to increase in accordance to Moore's law. The Internet is enabling cloud computing, making enormous amounts of computation available even to the smallest devices, as long as they are part of the cloud. The idea of crowdsourcing even leverages human computation for technological artifacts. It seems only to be a matter of time before the amount of computation available to robots exceeds the computational capacity of the human brain. Already today, cellular phones can connect to server farms to recognize spoken commands. And it is obvious that this is only the beginning ...

It is not easy to compare the computational capabilities of silicon-based computation with those of the human brain but an attempt might be interesting. A super computer consumes megawatts of power (costing millions of dollars to operate per year), the human brain roughly 20W. It is probably impossible to perform a reasonable comparison of computational power. The fastest supercomputers today perform in the petaFLOP range, performing thousands of MFLOPS per Watt. It does not make much sense to use FLOPS to measure the computational performance of the human brain, as humans are slow at performing floating point operations. However, it would be fair to say that brains today can do many things that supercomputers cannot do and that they are many orders of magnitude more energy-efficient than supercomputers. Computation has caused much progress in robotics, no doubt. And it will continue to do so. But there are three factors we should consider before we place our bets on computation as the winning paradigm.

First, Moravec's paradox states that high-level reasoning seems to require relatively little computational resources, whereas sensorimotor skills on the other hand seem to require huge amounts. Sensorimotor computation, however, appears to be fundamentally different from arithmetic as it connects the external world and the internal world of an agent. Aspects of uncertainty, representation, feedback, memory, learning, adaptation, etc. become a central factor, not so much pure number crunching. So it might be necessary to develop a deeper understanding of sensorimotor computation before we can leverage arithmetic computation in that domain.

Second, there is a theoretical limit to how much computation we can perform on earth; it is called the Bremermann's limit and is based on Einstein's mass-energy equivalency and Heisenberg's uncertainty principle. A computer the size of the earth, according to this limit, could perform 10^{75} FLOPS per second. We will never reach this limit but this number serves as a reminder that providing computation comes at a cost. Does it make sense to give millions or billions of robots access to supercomputers when there is a well-working example of distributed, low-energy computation in the human? This brings us to the final point: energy.

A back-of-the-envelope estimate from the year 2007 shows that computers and the Internet consume over 5% of world's total energy production [9]. This number may not be accurate but it illustrates that autonomous and intelligent robots have to be economically viable and therefore energy-efficient if they are ever to leave academic laboratories and factory floors.

1.4.3 Machine Learning

Machine learning is maybe the greatest success story of the last decade and therefore an obvious candidate for a winning paradigm. However, there are convincing arguments that machine learning by itself will not be able to solve the "pixels to torques"-problem. Taking the "pixels" as source of input literally, the number of different ways to arrange the pixels is so large, that methods capable of bootstrapping the "pixels to torques"-problem ab initio would also be able to break modern encryption algorithms [2]. Given the complexity of problems addressed by state-of-the-art machine learning techniques today, this point seems nowhere near.

The successes of machine learning have taught us a valuable lesson though: many interesting problems can only be solved if one provides an appropriate bias, i.e. information about the inherent structure of the problem, encoded in such a way that the learning algorithm can leverage it to guide the learning. In good old-fashioned AI such biases were called heuristics.

And this is an important insight to remember when thinking about new paradigms: many problems, when formulated in a naïve way, are too high-dimensional to be solved by brute force computation. And I would claim they will remain too high-dimensional, even if we wait for Moore's law to play out for another couple of decades (see discussion of Computation above). It is not a question of considering robots with more degrees of freedom, i.e. 50 instead of six. And it is not even a question of including the degrees of freedom of objects manipulatable by the robot – now we are talking about thousands, maybe millions dimensions to consider. A naïve formulation of the problem of autonomous, intelligent robots must include also all information available to the robot. The robot has to consider what it knows to change its environment so as to achieve a task. Now we are talking about billions of dimensions.

Personally, I think this is a strong argument in support of needing appropriate, i.e. problem-specific biases. More computation and smarter (but unbiased) machine learning methods will not be able to solve the “pixels to torques”-problem. Finding good biases might not be an easy problem but it is probably easier than overcoming the limitations of computation.

1.4.4 Morphological Computation

Morphological computation [6] captures the idea that morphology itself can perform computation. For example, a hand closing around a rock is “computing” an optimal grasp through compliance in skin, muscle, and tendons. This kind of computation is particularly well suited to address uncertainty in the world and therefore could form a good basis for grounding appropriate higher-level symbols.

The morphological computation paradigm implies that the sensorimotor capabilities of a robot should not be divided into hardware and software. Instead, computation must be distributed across these two components in a favorable way. In most of robotics this insight has at most been exploited accidentally. Is it maybe part of an explanation of Moravec’s paradox? Biological intelligence relies on morphological computation extensively. And maybe a winning paradigm for intelligent technological artifacts must also exploit the synergy of hardware and software in more direct ways.

1.4.5 Sensor Technology

The SICK laser range finder has revolutionized mobile robotics; it provided a huge improvement over sonar sensors. This improvement made finding an elegant and tractable solution to the SLAM problem possible. Recursive estimation and probabilistic robotics are now part of the standard tool box of any roboticist.

It would not be unreasonable to think that a similar improvement in 3-D sensing can boost the state of the art in robotic manipulation. However, this is based on the assumption that acquisition of information about the world is the bottle neck. If one considers seriously the implications of Moravec’s paradox and believes in the fundamental idea behind morphological computation, one might be led to believe that this is not enough. While better sense information will necessarily lead to improvements, the overall problem of manipulation at human dexterity-levels might require not only better sensors but also approaches that have an integrated perspective of sensing, perception, actuation, and action.

1.4.6 Development

Piaget's theory of stage-wise development in humans remains widely accepted in developmental psychology, 60 years after its publication [5]. An increasing amount of evidence is uncovered by neuroscientists, psychophysicists, and psychologists that even basic cognitive functions, such as those of the visual system, are learned incrementally by building on functions learned earlier [1]. It would therefore appear that stage-wise development plays a pivotal role in the benchmark system for autonomy and intelligence – the human.

The young and emerging field of developmental robotics (or epigenetic robotics) attempts to build autonomous robotic systems by following the idea of development. All too often, work in this field is concerned with learning about developmental patterns observed in humans and re-implementing them on robots. The mechanical and computational substrate used for these experiments is rather general and it is therefore no surprise that developmental patterns can be imitated and replicated by careful system design. The reimplementation of developmental phenomena is only interesting if it generates new observations that can be verified in the biological system afterwards. This is rarely the case in developmental robotics.

The more pertinent question may turn out to be: what are the principles of development in biological intelligence irrespective of the substrate or a particular developmental pattern that can be formalized and leveraged to build better robots? It seems that those principles must exist, as nature seems to have evolved several architectures exploiting the same principles [3].

One of those principles, so it would seem, must relate to compositionality. The incremental and stage-wise learning of cognitive capabilities in humans and animals realizes increasingly complex skills based on simpler ones. Good building blocks are therefore characterized by how well they enable more complex skills and not by how well they compose into a system that solves a particular problem, as it is the case in the system engineering paradigm.

The same principle should also relate to the aforementioned biases. One could imagine that the acquisition of a skill is only possible in the presence of particular biases that structure the input/output domain of the agent. Skills, once acquired, provide a different structuring of this space, in turn enabling the acquisition of other, more advanced skills. Compositionality and biases could therefore be interdependent.

Human development relies on the genetic program we are born with. The nervous system contains hard-wired processing elements that structure the raw data sensed by the 300 million external receptive nerve endings in the human body. We know that this structuring of raw sense data provides a kind of processing and behavioral basis without which development would not be possible. It is therefore likely that development in technological artifacts must rely on a similar computational and behavioral basis. Hence, development might be a part of a winning new paradigm but probably not the whole story.

1.5 What's the new paradigm?

Only the future can tell if a paradigm shift indeed is needed in robotics. In the brief discussion above, no obvious winning paradigm emerged. Maybe a new, yet unknown paradigm will take center stage in the coming years. We can only speculate about the path that leads us there. Here is my personal speculation:

- Our field should de-emphasize application-specific research. There are too many fundamental problems that remain unsolved (manipulation, segmentation, object recognition, etc.); our community should explore novel ways to deal with these problems, free from the pressure to produce "products."
- Research should consciously disregard the existing boundaries of sub-disciplines in robotics, for example between computer vision, manipulation, planning, control, and mechanism design. Instead, our field should focus on the development of truly integrated sensorimotor systems. This differs from the systems engineering paradigm in that the **development** of the system includes these aspects; in the system engineering approach the resulting system integrates these aspects but the aspects have been developed independently.
- Researchers should pay particular attention to the compositionality of skills. The quality of a skill cannot be determined by evaluating the skill by itself. The quality of a skill also encompasses the degree to which a skill enables or facilitates the performance of other skills. For this type of research we must move away from an experimental methodology that tests skills in isolation and instead move towards testing skills in the context of larger systems. This implies that skills have two equally important functions: the skill itself and also a structuring and simplification of the overall state space that enables more complex skills to be realized.
- The field of machine learning should make the study of appropriate biases a core part of its research agenda. We need to understand how task-specific biases can be identified, extracted, represented, and leveraged in learning.
- If we believe that development and autonomous learning are important avenues towards autonomous, intelligent robots, then it might be advisable to focus on a set of basic sensorimotor skills that provide sufficient structures on the "pixels to torques"-problem space. This might be viewed as the computational and behavioral

basis provided to biological intelligence in the form of genetically coded hardware skills.

References

[1] R. P. Würtz C. Prodöhl and C. Von Der Malsburg. “Learning the gestalt rule of collinearity from object motion”. In: *Neural Computation* 15.8 (2003), pp. 1865–1896.

[2] A. Censi and R. M. Murray. “Uncertain semantics, representation nuisances, and necessary invariance properties of bootstrapping agents”. In: First Joint IEEE International Conference on Developmental Learning and Epigenetic Robotics, 2011.

[3] N. J. Emery. “The Mentality of Crows: Convergent Evolution of Intelligence in Corvids and Apes”. In: *Science* 306.5703 (2004), pp. 1903–1907.

[4] T. S. Kuhn. *The Structure of Scientific Revolutions*. 1962.

[5] J. Piaget. *The Origins of Intelligence in Children*. 1952.

[6] C. Scheier R. Pfeifer and A. Riegler. *Understanding Intelligence*. 2001.

[7] C. A. Rosen and N. J. Nilsson. *Application of Intelligent Automata to Reconnaissance*. 1966.

[8] Unit E5 Cognitive Systems. *FP7-ICT CALL 6 - Questions and Answers*. 2009.

[9] `http://uclue.com/?xq=724`.

2 The Well-Rounded Roboticist

Gaurav Sukhatme, Stergios Roumeliotis | *U. of Southern California, U. of Minnesota*

Technical societies are often identified by the common background of their members usually acquired through undergraduate and graduate training. Members of some fields, despite the differences in the curricula of different programs among schools across the world, exhibit a degree of common background. As an example, it is safe to assume that the vast majority of the members of the IEEE Controls society have taken courses in control, and a random attendee at the CDC can correctly define a Lyapunov function.

This is not obvious for robotics where it is difficult to determine the defining core of our community in terms of common technical background. Various factors contribute to this. Most importantly the inherent diversity of the discipline means that robotics programs are housed in many different departments (e.g., Computer Science, Electrical-, Mechanical-, Aerospace Engineering, etc.) – probably a unique characteristic of our discipline.

While this pluralism enriches our society, it sometimes also hinders communication since people with different training often fail to understand and appreciate different viewpoints on the same subject, thus missing the opportunities that result from synthesis of different ideas. One often sees this e.g., at program committee meetings in the arguments between area chairs from different areas within robotics.

2.1 Introduction

The main question that poses itself is: *Is a common core for robotics possible to envisage, desirable to implement, and achievable in reality?*

2.1.1 Background

In the US[1], with one exception (Carnegie Mellon University), all PhD students working in robotics research are housed within the confines of a degree granting department. For most students, as far as their universities are concerned, the students are working on a degree in compter science or mechanical engineering or electrical engineering and *specializing* in robotics. The coursework the student takes emphasizes the broader departmental discipline. This can mean that coursework in robotics is marginalized. This creates a silo effect and the exposure of robotics students to the full breadth of topics

[1] The authors are not sufficiently familiar with graduate education outside the US so the examples given here are from the US only.

in robotics is limited. Large interdepartmental labs (e.g., GRASP at the University of Pennsylvania) blunt the impact of this effect to a certain extent.

The Carnegie Mellon University offers a true robotics PhD program whereas the Georgia Institute of Technology and the University of Utah offer coordinated “robotics tracks” for their PhD students housed in multiple departments. At Georgia students who want to specialize in robotics are advised to take two linkage courses on “Multidisciplinary Robotics Research”. The remaining coursework is department specific. At Utah, students in Mechanical Engineering and Computer Science can specialize in “robotics tracks” with differing requirements and a common core. Other universities (e.g., the University of Southern California) offer similar track arrangements but only in limited departments (typically one or two out of CS/EE/ME/AE[2]). The coordination between a PhD robotics track in CS and a robotics track in EE or ME is poor.

At the Carnegie Mellon University, the robotics PhD program requires students to take four core courses, one from each area: Perception, Cognition, Action, and Math Foundations. Perception core courses include “Computer Vision” and “Sensing and Sensors”, Cognition core courses include “AI” and “Machine Learning”, Action core courses include “Kinematics, Dynamics, and Control” and “Mechanics of Manipulation” and the Math foundations core course is a specially designed course for roboticists called “Mathematical Fundamentals for Robotics”.

2.1.2 What should the well-rounded roboticist know?

To seed discussion, we list broad topics that we believe robotics students should be familiar with:

- Kinematics and dynamics e.g., *Siciliano*
- Mechanisms and actuators
- Linear control e.g., *Hespanha*
- Sensors and sensing systems
- Motion planning e.g., *LaValle*
- Machine vision e.g., *Forsyth and Ponce*
- Artificial intelligence e.g., *Russell and Norvig*
- Machine learning e.g., *Mitchell*

[2] **C**omputer **S**cience, **E**lectrical **E**ngineering, **M**echanical **E**ngineering, **A**erospace **E**ngineering

- Statistics
- Linear algebra e.g., *Strang*
- Optimization e.g., *Boyd*
- Estimation
- ...

Given that there is a huge selection of topics and most of us would like students to learn to do research there are two choices: Either we accommodate many of the above topics into a set of coherent courses and campaign at our institutions to train roboticists or we decide that a common core is not important and teach a selected subset of the above to students depending on their (and our) departmental affiliations.

2.1.3 Food for thought

The authors advocate the former proposal – a specialized robotics PhD. Open questions are: Is there a rational middle ground? What is the European view on this topic? Where or in which cases is formal coursework not part of the PhD program? If students were trained more broadly would they produce better work? Be more creative? Would the quality of the reviewing process in our conferences and journals improve? Could we play a more effective role in our various national lobbying organizations by graduating broadly educated PhDs who share a common foundation? Would robotics technology commercialization speed up?

2.2 Meeting Discussion Summary and Outcomes

2.2.1 Questions and Discussion

- Should there be a specialized PhD program in robotics? No consensus was reached on this point.
- Wouldn't defining a core educational background for roboticists be useful? There was some agreement on this topic. Defining the field clearly to professional organizations (e.g., industry, government) could ultimately help academic institutions.
- Should there be a core set of subjects for robotics? If yes, who should define this core? The sense of the meeting was that the field is too comprehensive to define a core.

- What role do Wikipedia, Google and online lectures play in education? Some felt that we need short modules to be available on demand to all students, rather insisting on a core set of educational requirements.
- The point was made that robotics needs strong links between lectures and labs. Shouldn't this be part of the core?
- How do we define the core? Experts certainly have opinions. Should it be the union of topics defined by us? Or the intersection of topics defined by us? There was no consensus on this topic.

2.2.2 List of Topics Suggested by Attendees

Each of these topics represents half-a-semester course at a US academic institution:

- Kinematics
- Dynamic systems
- Filtering and estimation
- Math
- Linear algebra
- Differential equations
- Probability theory
- Optimization
- Inference
- Adaptative control
- Motion planning
- Navigation and mapping
- Sensing systems
- Power electronics
- Vision
- Digital systems

- Algorithms
- Mechatronics
- Machine learning
- Mechanics/mechanical design
- Programming
- Graph theory
- Logic
- Planning (distinct from motion planning)

2.2.3 Next Steps

A subcommittee is being constituted (chair: Sukhatme) to generate a document that will have a list of important topics (candidates for the list are above) and:

- A short introduction (1 paragraph) to the topic
- Relationship to other topics
- References to monographs/articles for practitioners
- References to books for fundamental understanding

3 Are we even in the game?

Vijay Kumar, Matt Mason | *U. of Pennsylvania, Carnegie Mellon University*

We may feel that robotics research has made enormous strides over the past 50 years, but the real world impact is disappointing to some. How far have we come, and how far is left to go? This document summarizes several email exchanges before the Berlin Summit and at the Summit. We hope this will continue to provoke discussion on the current state and future direction of robotics research.

3.1 Introduction

In 1961, there was Unimate, and Joe Engelberger appeared on "The Tonight Show" with Johnny Carson and showed the Unimate perfectly putting a ball to loud cheers. The same software presumably was used to handle hot ingots on the shop floor. On December 23 that same year, Heinrich Ernst defended his doctoral thesis (supervised by Claude Shannon) in which he described the MH-1, a computer-operated mechanical hand (Fig. 3.1). He developed programs that would, for example, "search the table for a box, remember its position, search the table for blocks, take them and put them into the box." As he says in his thesis, "The position of the objects is irrelevant, as long as they are on the table." His "previous research" includes ideas of building a versatile hand attributed to a seminar by Claude Shannon and Marvin Minsky. (A film of Ernst's project is available online at `http://projects.csail.mit.edu/video/history/aifilms/15-robot.mp4` and an accompanying narration, recorded later, is at `http://projects.csail.mit.edu/films/aifilms/Podcasts/15-robot.mp3+`

We roboticists are trying to figure out how to build mechanical people. How far have we come? How far do we have yet to go? We have made a lot of progress. Our computers are certainly faster (Ernst's choices were between a TX-O and an IBM 709). We have power programming abstractions and simulation tools. Hydraulics and pneumatics with simple on-off actuators are now replaced by backdriveable brush-less DC motors. Certainly our arms have incredible precision compared to the foundry manipulator. We have planning algorithms that can explore 11-dimensional spaces. Clearly we have made progress. Today's graduate student is incredibly sophisticated in her approach to solving Ernst's problems and Ernst would be envious of our research laboratory.

Are we nearing a finish line? If so which finish line is it, and how far is this finish line. Even though we have made a lot of progress, what is the progress since Ernst's work with respect to the finish line? Related to this is a fundamental difficulty. We can measure degrees of freedom, speed, force, precision, accuracy, of the robotic mechanism, but our measures of the actual task performance are almost non-existent. How would we even construct a numerical measure with which we could mark our progress or compare competing systems? Can we even say what capabilities ultimately define a general-purpose manipulation system? How do we define a true dexterity index? Given that we do not have answers to these questions, it's hard to imagine we will approach a goal within the next 50 years!

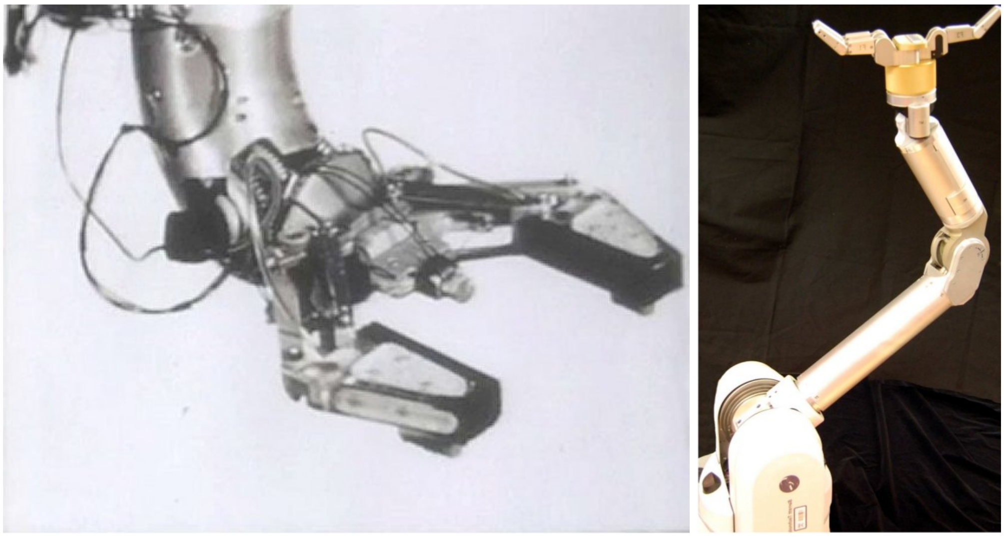

Figure 3.1: Left: Ernst's 1961 computer-operated mechanical hand, MH-1, which was attached to an off-the-shelf American Machine and Foundry Servo Manipulator; Right: Barret Technology's 2011 four degree-of-freedom hand attached to a seven degree-of-freedom manipulator.

3.2 Accomplishments

3.2.1 Impact of contributions from the robotics community

This is a compilation of things we can be proud of and have made a real-world impact:

- Work on manipulator kinematics, manipulator design, and manipulator programming systems has been incorporated into all industrial robot systems.
- We can model the three-dimensional geometry of complex environments and reason about configuration spaces in high-dimensions and plan reaching movements that avoid collisions and joint limits.

- Some advanced control algorithms have also made it into industrial robot arms - the computed torque algorithm, a certain class of adaptive control algorithms, and force control algorithms.
- Robot architecture, control, and even autonomous navigation software runs on Spirit and Opportunity on Mars.
- Some automobiles can parallel park autonomously.
- Our mechanical design, control, and teleoperation methods are deployed in surgical systems.
- Simple perceptual, control, and navigation systems are used in vacuuming robots. (Some of them do navigate, for example the Neato Robotics and Evolution Robotics products.)
- Autonomous systems are making inroads in shipping yards, mining, agriculture, and unmanned air vehicles.

3.2.2 Significant contributions that may not have had the impact we once thought

Here is another list of accomplishments. Have these accomplishments made a difference to grasping and dexterous manipulation in unstructured environments?

- Grasp synthesis and analysis – mathematics and algorithms of caging, immobilizing and grasping ensuring stability – many impressive, well-cited papers but how useful has any of this proved today?
- Other manipulation – folding towels and origami, catching a tumbling cell phone in a fingertip grasp, throwing things, even juggling . . . yet these point solutions show now sign of approaching a more general competency.
- Learning – very impressive appearance-based learning, imitation learning, reinforcement learning - nothing that appears to generalize across even a narrowly-defined set of tasks (how to define this is also a problem).
- Perception – incredible progress in object recognition, image-understanding . . . and yet, we are nowhere close to achieving real-time perception-action loops. The vision folks are not interested, perhaps because the problems are too "dirty" (and way too hard!).

- Haptics – models and demonstrations of hands feeling and recognizing objects – presumably a lot of potential but few practical solutions for grasping and manipulation.

Even in automated factories, humans are making a comeback. Smart phones are generally assembled by hand. In some other cases, automated assembly operations have fallen into disrepair and been replaced by humans.

3.3 The DARPA ARM-S project

Robotics has been exposed! Witness the tasks in figure 3.2 that top researchers in the US had to do in a *very* structured environment (table top, good lighting, perfectly positioned stereo rig and ranging sensor) with a state-of-the-art manipulator and arm. Six teams competed. The best team completed task 1 four of five times with an average completion time of 136.1 seconds. The best team completed task 2 four out of five times with an average completion time of 153.7 seconds. The best teams completed task 3 two out of five times with average completion times of 84 and 274 seconds. Three of six teams either did not attempt or complete this task. After 50 years, shouldn't these tasks be homework problems in a robotics undergraduate class? Arguably the simple environment in figure 3.2 is a structured environment and industrial robots have been used successfully in such environments. Why is this task so difficult?

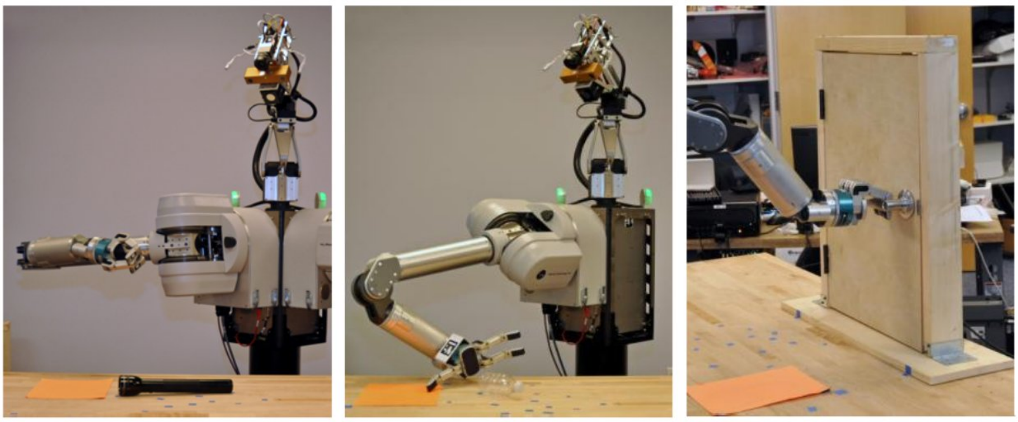

Figure 3.2: Challenge tasks in the ARM-S Project. Locating, grasp and placing a flashlight (left: Task 1) and a bottle (center: Task 2) on the orange paper and opening a door (right: Task 3)

3.4 Measuring Unstructuredness

One of the greatest weaknesses of our field is the inability to precisely characterize a task. When measuring progress we have no basis for comparing across task domains. Is an ant's job more difficult than a human's? Surely not, but how do we make that precise? Are today's robots "better" than ants? We cannot answer this question today. We don't have any precise way of characterizing different tasks or task domains.

One commonly applied characterization of a task is structured versus unstructured environments. Unfortunately the terms are used without consistency or precision. We talk of the kitchen as an unstructured environment, but we really depend on having a flat floor. Is there a precise way of comparing the uncertainty or structuredness of a factory with a kitchen? It is important that we find a way. Some people feel that factories are fundamentally uninteresting, that they are too structured, that there is no uncertainty, and hence that nothing is to be learned from experience in factory automation. A more precise understanding of structured and unstructured environments might enable robotics researchers to integrate lessons learned in factory automation.

3.4.1 History of "unstructured environments"

A quick literature search turned up about 200 papers mentioning unstructured environments, most of them only in passing. Most papers use the term without a definition. Surprisingly, the term was in use as early as 1967, used in SAIL Memo 51, by Lester Earnest. His observation was that in order for a computer to operate efficiently in an unstructured environment, it has to have something like an eye.

Various subdisciplines of robotics use the term in various ways. In navigation, it sometimes means that the robot has a map. Sometimes it means that the road has curbs. To some people it means that beacons were installed. Shakey was a mobile manipulator with significant uncertainty in the locations of objects, but later researchers call it a structured environment. Only certain things were to be found in the environment, they were very regular in their appearance, and features such as visually contrasting baseboards were added to the environment to aid localization.

3.4.2 Definitions

In manipulation research at least, it seems the dominant idea is that if you design both the robot and the task environment, that's a structured environment. Factories are the canonical example. Let's call that the "co-engineering" definition.

Definition: A structured environment is an environment that is co-engineered with the robot for solution of the task.

Examples of the co-engineering definition of structured environments would be beacons in a navigation task, or parts feeders in a factory. What happens if an environment was structured for a *different* robot? Or, more generally, for a different agent, including a human? There isn't a clean line of separation. Imagine you're in a factory, and you say this is an unstructured environment. And the engineer says, well it used to be, but then the robot broke and we swapped in a new robot. So, we introduce the "engineering" definition:

Definition: A structured environment is engineered for solution of the task.

An example of the engineering definition of structured environments would be the kitchen, which is engineered for ease of human solution, but with regularities such as flat floors that also ease robotic solution. A third possibility is that the environment was not structured for the task, but it has regularities that the robot can exploit. So, we introduce the "regularities" definition:

Definition: A structured environment has regularities that simplify a task.

An example of the regularities definition would be the curbs on the side of a road, which are actually present to deal with rainfall (I think) but which can be exploited by range sensors to find the road edge. A cynical view of the robotics literature suggests an additional definition:

Definition: A structured environment is the environment solved by the previous researchers.

Each time we approach a new task, the environment is frighteningly open and unstructured. We ultimately solve the task by identifying and exploiting regularities, and perhaps introducing some regularities. Having accomplished that, the regularities are very obvious to future researchers who now perceive it as a structured environment.

3.4.3 Implications

The real issue is, what expectations does the robot require of the task environment for it to do its job? If you are designing a bot for a kitchen, you should optimize for counters of standard height, for the utensils and appliances commonly used by humans. Drinking glasses are not made of soap. Kitchen floors are not cobblestone or sand.

Some implications of these definitions:

- Structuredness is not binary. Whatever definition one chooses, structuredness varies over a continuum.
- Structuredness is relative to the task. If the task is to emulate a statue, many uncertainties in the environment might be ignored.
- Structuredness is relative to the robot.

We stipulate some additional observations:

- All task environments have some regularity, derived from whatever source.
- If you know of these regularities, and can exploit them in the design of the robot, you should.
- If you can add to that regularity offline, inexpensively, you should.
- If the robot or human wants to add to that regularity online, it can. That's an essential part of many big tasks, especially kitchen work – creating and maintaining order.

We still lack a means of measuring structuredness, but perhaps we have made some progress. The idea would be to measure that order in an environment that facilities solution of the task for the given robot. If we could agree on a precise metric, we could then go back to Les Earnest's characterization. Is there actually an important distinction between uncertainties typically faced by a factory robot and those faced in the kitchen. Factory robots face rampant uncertainty, but generally they don't need eyes. Typically they deal with uncertainty with a few binary sensors, plus mechanical compliance, and perhaps even some tactile sensing and force sensing. But it is essential that they know *approximately* where things are.

Tying this back to earlier sections, we might look at the variety of demos/videos by our community that claim progress. The PR-2 community has collectively bagged groceries, folded towels and opened/closed doors. Andrew Ng and coworkers learn object orientation and successful grasp strategies from appearance. Srinivasa has very impressive demonstrations of an arm reaching and grasping in cluttered three-dimensional environments. Hsiao, Kaelbling and Lozano-Perez have used tactile sensing to grasp and manipulate a variety of objects with uncertainty showing over 90% success rate.

And yet we have the embarrassing DARPA ARM-S experience. Could it be that the impressive demo's are nothing more than proof that the line between "structured" and "unstructured" environments is malleable? Should we interpret the difficulty of the ARM-S challenges as evidence that they are well-conceived, or poorly conceived?

Is it time for us to organize well-defined challenges so as to measure progress and encourage efforts in well-motivated directions? Or, is our field still so new that it makes sense for us to continue the loosely-organized-stabs-in-the-dark approach?

3.5 Possible challenge tasks

Some robotic challenges have been very successful – witness the DARPA Grand and Urban Challenges for autonomous driving, and the robot soccer competitions. There are also competitions in Search and Rescue. What would be good challenge tasks in manipulation?

3.5.1 Personal/service robotics challenges

Several groups are working in kitchens. Should we devise challenges or competitions related to kitchen work? Loading the dishwasher? Extracting items from cabinets and refrigerators? Unjamming the garbage disposal? Brushing the ants off the dessert without the guests noticing? It raises some interesting questions in defining not just one set task environment, but a distribution of task environments, or perhaps a continuum of distributions of graded complexity.

Towel folding races? Gathering coke cans? What about traditional human games: Jenga, pick-up-sticks, tiddly-winks, sleight-of-hand? Actually, humans have *lots* of challenges and metrics and competitions for manipulation tasks. For a taste, check out `http://www.youtube.com/watch?v=3qjdYpKq4Yw` and `http://www.youtube.com/watch?v=Bu-Pj0OFyPU`. Interestingly, as far as we can find, humans do not have kitchen-cleaning contests, but they do have kitchen-messing-up contests.

3.5.2 Bin picking: A simple structured task

Bin picking is an interesting problem. There were lots of interesting projects on it, including some folks trying to automate the design of bowl feeders, or design general-purpose feeders, or even trying to do the most “obvious” thing: look in the bin, pick out an object, accurately determine its pose, plan a grasp, and go get it. Some notable efforts: Horn and Ikeuchi’s system for picking plastic donuts. And a recent commercial success story called Braintech. Successful, right up until it failed. Sigh. Still, it shows both interesting capabilities and limitations. A video is available at `http://www.youtube.com/watch?v=09Lzuf0nbX0`. Here’s an Adept video, somewhat similar: `http://www.youtube.com/watch?v=2zygwhgI03I`

Here's another more conventional automation approach to a simpler problem: `http://www.youtube.com/watch?v=wg8YYuLLoM0`. There was an interesting system called a Flex Feeder, developed by Adept. I think Adept gave up on it eventually, but I found this variation on it: `http://www.youtube.com/watch?v=v-Kh588NkVo`

It's a nifty set of recirculating conveyors that provide a randomly scattered stream of parts on an underlit stage. For some parts, it is enough to do 2D vision, find the parts that are correctly oriented and isolated, and go get them using preplanned grasps. You could imagine setting a set of challenges where you change the illumination, vary the types of parts, whether the parts are familiar or not, whether they are mixed, density of them, and measure performance as mean time to picked parts, precision with which a picked part's pose in the hand is known.

Some people don't like this stuff because it is "structured", and some would even say there is no uncertainty. Obviously we could have a long argument over that, but never mind. It is a task that lends itself to benchmarking and it is sufficiently complex that people would not compare it to what Ernst tried to do (and accomplished) in 1961.

4 Learning in Robotics: The complexity is in the environment

Marc Toussaint, Tomas Lozano-Perez | *FU Berlin, Massachusetts Institute of Technology*

4.1 Starting point

Learning in robotics has a long tradition that grew into a wide field. One way to categorize modern approaches to learning in robotics is in the formalization of Reinforcement Learning: Figure 4.1 provides such a taxonomy strongly biased by the RL formalism. When surveying such a spectrum of work there are several observations:

1. Most examples of successful Machine Learning (ML) in robotics concern control of the robot's own or attached DoFs. Only in relatively few cases has ML been used for a robot to learn how to control/manipulate a natural environment.

2. Integrated robotic systems are huge – in terms of lines of code, as well as mix of methodologies/disciplines/formalizations that comprise the system. Most examples of successful ML in robotics concern tightly constrained learning problems "sprinkled" throughout such integrated robotic systems. The engineer identifies a potential for learning, maybe on the control level, for object recognition, for calibration, or for automatic optimization/tuning of cost functions, etc. Nevertheless, the learning rarely concerns the system as a whole.

3. Tackling truly complex manipulation problems (many DoF and very long action sequences) requires exploiting/discovering appropriate hierarchies and abstractions. These are areas where progress in both ML and robotics has been slow.

We believe such issues are central obstacles for the progress of learning in robotics. While the naive dream of introducing learning to robotics (or AI in general) is that it makes systems more autonomous and will eventually spare us from pre-programming behaviors, this goal has – despite the successful demonstrations of learning on specific problems – not really been achieved. The question is why and what are promising research directions to overcome the mentioned obstacles.

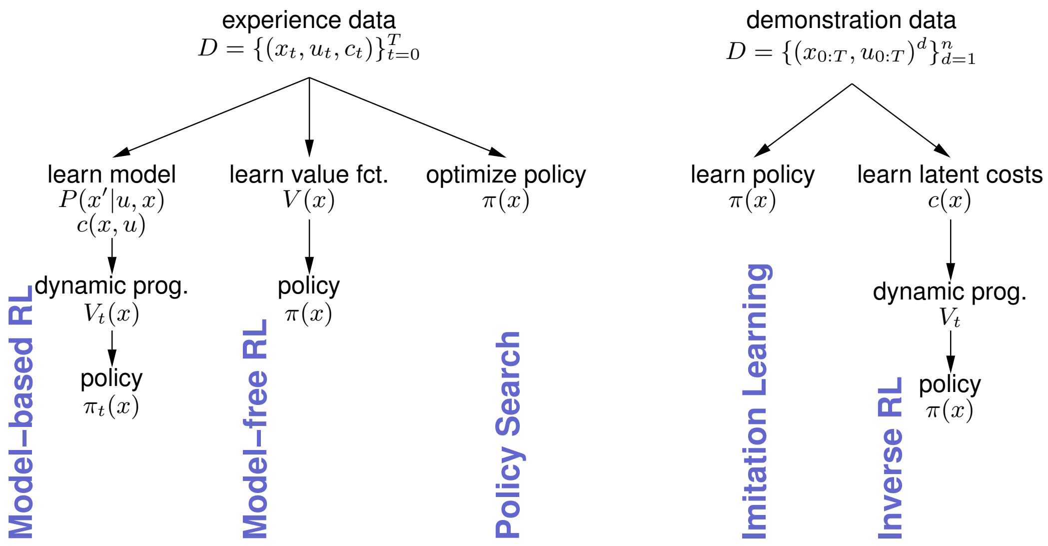

Figure 4.1: Five approaches to learning in robotics

4.2 Views on future challenges and premises for ML in robotics

During the discussion several views on the role of ML in robotics have been mentioned. To reflect this variety of views and opinions we briefly recapitulate some parts of this discussion.

4.2.1 Abundant data and longevity

Robotics in principle allows collecting huge amounts of data. One example mentioned in the discussions is navigation, where a system (e.g., autonomous car) navigating for years in real environments could collect data on an enormous variety of situations and conditions (Paul Newman). It is sometimes remarked that such abundance of data (combined with efficient retrieval methods) can make a qualitative jump in model performance, as demonstrated by current trends in image analysis and internet applications. Following this line of argumentation we should aim at longevity and life-long learning of robots, and on collecting more and more data.

From a more abstract point of view, this discussion is about the importance of the prior versus the data: On which should we put more emphasis? The data-oriented perspective diminishes the role of the prior, perhaps only relying on a basic nearest neighbor method for retrieval and lazy learning, in hopes that eventually we will have data for nearly any possible situation. In contrast, the prior oriented perspective asks which structural aspects of natural environments and types of generalization we should or need to incorporate in a prior to enable learning. The latter raises fundamental and open research questions on what specific priors in robotics could be, which were subject to further discussion as mentioned below.

4.2.2 The role of learning algorithms in robot architectures – towards the "Double Down sandwich"?

Some applications of learning algorithms in robotics play the role of filling some gap in a pre-structured control architecture. That is, in the given architecture there may be some mapping, classification or relation initially unspecified which is then estimated based on data. In the discussion this was coined a "sandwich". This is put in contrast to the "Double Down sandwich" perspective, where learning should address everything, including the architecture (Stefan Schaal).

It is interesting and perhaps controversial to discuss which existing learning paradigms should be viewed as "sandwich" vs "double down". All types of system identification, including learning motor dynamics and SLAM, could be viewed as just sandwich. The family of Reinforcement Learning algorithms could perhaps also be viewed as just sandwich, at least when the core RL algorithm is embedded in a larger system architecture. In fact, it is quite unclear how "double down" can actually work; how learning methods should be able to learn the architecture itself.

4.2.3 Focus on Robot dynamics vs. interaction and manipulation of environments.

Arguably, most successful applications of learning methods in robotics to date concerned "internal degrees of freedom" of the robot it self. By "internal" we want to include the robot position, as in navigation and SLAM, or directly attached DoFs like in balancing. This is meant in contrast to learning models of "external degrees of freedom" like how to manipulate objects in a natural environment. There are examples on the borderline, like learning to juggle, hitting a ball, etc. as well as learning models of interaction with one specific object. But this still is in contrast to learning higher level manipulation strategies in natural environments, where the focus is more on learning something about the environment than about the robot.

This issue seems related to the reduced emphasis on robot manipulation in the last decades of robotics research, where the success stories in motor control, motor skills, SLAM and motion planning, including learning methods in the these areas, have somewhat superseded the traditional focus on manipulation research.

4.2.4 Learning *for* reasoning – the non-decomposability of learning and reasoning.

A central question is *what* to learn. In some cases, e.g., standard system identification, it is fairly well-defined what the learning problem is because it is clear how the learnt model can then be used for control. In other cases, where planning and reasoning with given models is still an unsolved challenge, it becomes less clear what to learn. Generally, we should aim at learning models that enable efficient reasoning/motion planning/control. Therefore, in hard domains (like probabilistic relational domains) the research on learning methods cannot be independent from the research on reasoning methods using learnt models – only when we have an idea how reasoning could work efficiently we can have an idea what to learn.

To give an example: If the goal was a robot that "learns how to clean up a kitchen", it would be unrealistic to believe that two researchers could split the task: one taking care of the learning, the other of the reasoning. The first might do research on SVMs for some years, the other on STRIPS planning for some years, then the meet back and "simply put things together". Instead, both would continuously have to communicate what the learning constraints imply for reasoning and vice versa. For instance (arguably), the need for learning implies that reasoning must cope with probabilities. The need for reasoning in environments with many objects implies that learning must cope with relational structures. Generally, of crucial importance are the representations shared between learning and reasoning, whether they allow for both efficient learning and reasoning, and what priors they imply.

4.3 Promising directions

4.3.1 Logic, Probability and Geometry

Moving the focus of robot learning research beyond robot control and "internal degrees of freedom" towards learning higher-level models of the environment itself will have a series of implications:

While most learning tasks in the context of motor control, skill acquisition or SLAM are concerned with a fixed-dimensional vector space, the state space of the "outer environment" is exponential (e.g., in the number of objects) and naive, propositional vector-space representations are hardly promising. Therefore, simple off-the-shelf regression and classification methods are hardly useful to learn interesting models. The structures inherent in natural environments are very special and very strong – it will remain impossible to learn efficiently without translating this structure to more sophisticated representations and priors. *One* structural aspect of natural environments (out of many others) is that the effect of actions depends on relations and properties of objects, not their identity. This points to the research on *statistical relational learning*, which seems a good approach to reflect this prior structure.

Worlds of objects, characterized by their relations and properties, are classically described in AI by logical formalisms. These formalisms (relational representations, first order logic representations) can represent some of the underlying structure of natural environments, abstracting away from details of shape and location. However, the inability of logical formalisms to capture details of shape and location (geometry) led to the rejection of classical AI planning for robotics and to the development of geometry-based motion planning approaches. Recently, there has been progress in bridging this gap between logical and geometric representations, such as [4, 5, 12, 3]

But, *learning* (which includes modelling uncertainty, regularization, model selection) on these more complex representations has, until recently, been difficult. Recent progress of Statistical Relational Learning offers some hope that these problems can be tackled. The popular science article *I, algorithm: A new dawn for artificial intelligence* (Anil Ananthaswamy, NewScientist, January 2011 [1]) explains this field as "combining the logical underpinnings of the old AI with the power of statistics and probability." Stuart Russel is cited as "It's a natural unification of two of the most powerful theories that have been developed to understand the world and reason about it." The article goes on to suggest that with these new developments a new "AI spring" is arising after the "AI winter". Independently of how often this has been promised before, we think that at least this is going *exactly* in the right direction for robotics![2]

Our hope for SRL (enhanced with geometry) is that is has potential to get learning methods out of the bubble – to lift them to address learning generalizing models of the manipulability/controllability of worlds of objects based on their relations and properties. If it works, it will finally show that ML can really make a system more autonomous. The main current challenge is to ground and demonstrate existing learning methods in real robotic systems. Here is some work in this direction: [8, 9].

[1] `http://susansayler.wordpress.com/2011/02/01/i-algorithm-a-new-dawn-for-artificial-intelligence/`

[2] In fact, it leads back to the original root/motivation of Machine Learning as a discipline which should enable intelligent systems to interact with the environment. See Pat Langley's editorial on the origins of ML in the 25th Anniversary issue of *Machine Learning*: The changing science of machine learning `http://www.isle.org/~langley/papers/changes.mlj11.pdf`

4.3.2 Learning Hierarchy and Abstraction

Once we talk about "controlling" the environment (manipulation) and not just the robot, the size of the planning, control and learning problems becomes potentially unbounded. The only reasonable approach is to break up the problem into (hopefully small) pieces. Two powerful ideas that can help are temporal hierarchies (to break long sequences into shorter sequences) and abstraction (to focus on relevant aspects of the problem). This is an area with a long history in AI but characterized by relatively slow progress. It has received some recent attention within AI/ML/Robotics: [2, 10, 11, 7, 4, 5].

4.3.3 Learning on the system level

The discussion on (learning) system architecture raises the question: Can learning go beyond addressing isolated aspects (pre-determined by the engineer) and instead addressing the system as a whole.

An observation in this context is that ML generally works with a "full formalization" of the domain. Typical conference papers spend about a page to (very briefly) define the framework – of course by building on previous papers and formalisms, but being more or less concise and self-contained. What would be a proper formalization of an integrated robot system as a whole – including all its levels and components, not just focusing on one specific (e.g., symbolic or control) level? Characterizing the system abstractly as an MDP, POMDP or PSR (although correct) is not a sufficient answer. The model would have to reflect the overall structure of a robotic system, capture the approximations that make the problem tractable, and eventually enable learning within that whole structure.

Mitchell wrote a commentary on the *Discipline of Machine Learning* `www.cs.cmu.edu/~tom/pubs/MachineLearningTR.pdf` where he advocates as long term goal the "design of programming languages containing machine learning primitives". Indeed, some researchers [13, 1, 6] have attempted to design robot programming paradigms that inherently integrate learning, e.g., in which some subroutines are hand-coded while others are specified as "to be learned". Is this a promising way to go? Will this give the Double Down sandwich?

4.4 Concluding remark

While this discussion focussed primarily on learning, we belief that many of the conclusions would also hold w.r.t. the design of non-adaptive robots in the following sense: Even when we conceive of robotics aiming to engineer intelligent but non-adaptive systems, the process of designing such systems has analogous problems as processes of learning and adaptation. Development of a system based on inappropriate representations and structural decompositions will equally come to a halt like learning algorithms incapable to develop appropriate representations, hierarchies and abstractions. Perhaps in the field of software engineering such issues are most explicitly expressed, but such issues seem to penetrate all design, adaptation and learning processes.

The focus of discussion (and in particular this summary) had a strong bias towards addressing the structure of the environment. Promising research directions we pointed out are combining the concepts of logic, probability & geometry, Statistical Relational Learning as addressing two of these aspects.

However, the relational, logic and geometric structure of natural environments is only one of many aspects of natural environments. We mentioned geometry. Others are kinematic structures, structure that arises from Newtonian physics (e.g., that objects mostly do not float in the air), etc. When efficient robot learning (and reasoning) needs to exploit such priors, then robotics needs describe and uncover these structures.

One of the participants (Gregory Dudek) coined the expression “Robotics as the new Physics” (see 10). This can be interpreted in a educational and societal sense: Robotics will play a role as central as Physics in the past. Or it can be interpreted in the above sense that robotics becomes eventually a natural science which is, to a large degree, concerned with researching and describing the structure of natural environments. The latter corresponds to the shift of focus in robot learning and reasoning research on external degrees of freedom and higher-level models of the environment itself for robot manipulation.

The complexity is in the environment.

References

[1] D. Andre and S. Russell. “Programmable reinforcement learning agents”. In: *Proceedings of the 13th Conference on Neural Information Processing Systems (NIPS 2001)*. 2001, pp. 1019–1025.

[2] Jennifer Barry, Leslie Pack Kaelbling, and Tomas Lozano-Perez. “DetH*: Approximate Hierarchical Solution of Large Markov Decision Processes”. In: *International Joint Conference on Artificial Intelligence (IJCAI)*. 2011.

[3] Stephane Cambon, Rachid Alami, and Fabien Gravot. “A Hybrid Approach to Intricate Motion, Manipulation and Task Planning”. In: *International Journal of Robotics Research* 28 (2009).

[4] Leslie Pack Kaelbling and Tomas Lozano-Perez. “Hierarchical Planning in the Now”. In: *IEEE Conference on Robotics and Automation (ICRA)*. 2011.

[5] Leslie Pack Kaelbling and Tomas Lozano-Perez. “Pre-image backchaining in belief space for mobile manipulation”. In: *International Symposium on Robotics Research (ISRR)*. 2011.

[6] Alexandra Kirsch. “Robot learning language – integrating programming and learning for cognitive systems”. In: *Robotics and Autonomous Systems Journal* 57 (2009), pp. 943–954.

[7] G.D. Konidaris et al. “Autonomous Skill Acquisition on a Mobile Manipulator”. In: *Twenty-Fifth Conference on Artificial Intelligence (AAAI)*. 2011.

[8] Tobias Lang and Marc Toussaint. “Planning with Noisy Probabilistic Relational Rules”. In: *Journal of Artificial Intelligence Research* 39 (2010), pp. 1–49.

[9] Tobias Lang, Marc Toussaint, and Kristian Kersting. “Exploration in Relational Worlds”. In: *Proc. of the European Conf. on Machine Learning (ECML 2010)*. 2010.

[10] Bhaskara Marthi, Stuart Russell, and Jason Wolfe. “Angelic Semantics for High-Level Actions”. In: *ICAPS*. 2007.

[11] N. Mehta et al. “Automatic Discovery and Transfer of MAXQ Hierarchies”. In: *International Conference on Machine Learning (ICML)*. 2008.

[12] Moritz Tenorth et al. “KNOWROB-MAP – Knowledge-Linked Semantic Object Maps”. In: *Proceedings of 2010 IEEE-RAS International Conference on Humanoid Robots*. 2010.

[13] S. Thrun. “Towards programming tools for robots that integrate probabilistic computation and learning”. In: *Proceedings of the IEEE Int. Conf. on Robotics and Automation (ICRA 2000)*. 2000.

5 Perception-Action-Learning

Stefan Schaal | *U. of Southern California*

It appears that the largely monolithic approaches to perception, control, and learning have reached their limits and a more integrated approach is needed. Interestingly, if one examines ICCV statistics, one will find hardly any papers on perception for manipulation. Active vision, i.e. perception that includes moving sensors, seemingly has been disappeared since many years ago, despite it was quite popular in the 1990s. Similarly, how many robotics researchers actually work with real perception systems, and how many papers can be found where people address control strategies to improve the quality of perception? Again, there is not too much out there. Learning adds another component to these problems. There is a community that cares about learning for control, although the number of people working on machine learning for complex robots is rather small. Obviously, machine learning is part of computer vision, but there is not too much work where people try to devise strategies how perception systems learn competency in a bottom-up approach, e.g., the idea of "autonomous perception systems".

And perception also includes tactile perception and acoustic perception, which are rarely addressed in robotics research. Naturally, sensor fusion becomes important in multi-modal perception, a topic that researches in mobile robotics and state estimation have looked at, but that has not found wide spread attention in robotics systems, particularly when fusing vision, haptics, and audition. In contrast, the importance of perception-action cycles has been emphasized in psychology for a long time. Multi-modal and cross-modal perception is an upcoming topic in cognitive science. Thus, a bigger question for robotics becomes how to start a more comprehensive approach to perception-action-learning systems, an approach that emphasizes the need to address all these topics in an integrated way rather than treating them as independent research topics.

5.1 Introduction

A common feature shared by most autonomous systems is the concept of a perception-action loop. Biological systems interact with the world by perceiving relevant aspects of the world and their own state, processing these aspects, and continuously deciding what actions to take based upon this information. While the computations involved in the processing of perceptual data and generation of actions – and in particular the closed-loop properties of perception-action systems – are far from understood in biological systems, it is clear that they involve aspects of learning and adaptation as well as processes of inference in the face of uncertain and hidden information derived from very high dimensional multi-modal data streams. Moreover, perception, action, and learning are tightly linked into a functional system, i.e. it is hardly conceivable that the building blocks were developed independently of each other.

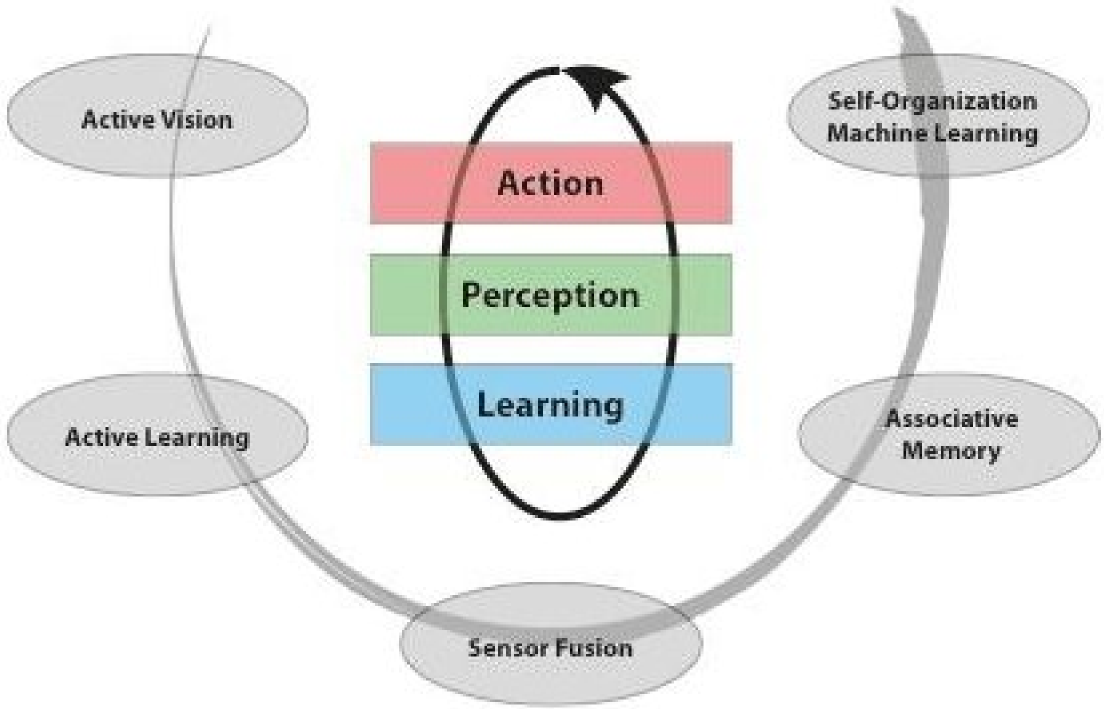

Figure 5.1: Perception-Action-Learning Loop with some relevant research topics, arranged from simpler to more complex topics (left to right)

Synthetic autonomous systems interacting with the world fundamentally face the same challenges; they need to perform actions which, based on noisy measurements taken in a complex world, let them carry out given tasks; e.g., maintain their structural integrity (homeostasis), navigate in an environment, manipulate an object, or harvest energy. While in the 1980s and 1990s, robots were largely dominated by minimal sensing capabilities and most emphasis was given to issues of control and planning in close to deterministic and static environments, the advent of cheap 2D and 3D vision and the push for haptic and auditory perception is in the process of changing this picture, an issue even more emphasized by the desire to take robots into human domains which are inherently dynamic and stochastic. It becomes apparent that trying to understand autonomous robots primarily from the viewpoint of control and planning is going to fail, and trying to approach perception without taking into account what matters for control/planning and how control can help perception is equally a doomed approach. Given the complexity of modern complex robots, the large amount of data available, and the increasing number of sensors, machine learning, i.e., automatic and adaptive tools for data interpretation, seems to be a key component in future robotics.

Thus, studying complete perception-action-learning loops (Figure 5.1), not in a component-wise isolation but rather as an integrated system will most likely form a new focus of robotics research. Then, the bigger question becomes, how to create a research methodology that aims at generating a general understanding of perception-action-learning, rather than specific non-general case studies. The following issues seem to play important roles in learning in perception-action loops, although none of them by itself answers the question of how to generate general perception-action-learning systems.

5.1.1 Active Perception

Active vision, i.e., vision with moving cameras inspired by the oculomotor system of humans and animals, was a popular research topic in the 1990s [1]. In the last 10 years, it has seemingly dropped out of favor – some vision colleagues claim that supporters of active vision were too dogmatic and did not deliver exciting enough results. One component that has survived is research on visual attention [12, 13, 11], although the groups that work with moving cameras often do not overlap with the groups that work on attention algorithms, which can be done solely based on video streams from static cameras. Some researchers consider visual attention as the highest level of active vision [11].

Using a motor system to improve perception can obviously be more general. Moving a head to improve auditory perception, or moving an arm with a touch-sensor-equipped hand can provide improved information about the environment (e.g., [7]). Robot manipulation can also re-arrange objects for better perception in various sensory modalities, or to identify additional properties of an object, e.g., internal degrees of freedom (e.g., [8]). The navigation and state estimation community (e.g., [16]) has also contributed to active perception by means of mobile sensors on a robot, and research on sensor networks (e.g., [3]) is often a form of active perception, e.g., in order to optimally extract information in a complex domain with minimal resources.

5.1.2 Active Learning

Closely coupled to active perception is the topic of active learning, i.e. how to choose where to sample new data points in order to optimize information gain. Again, this topic was popular in the 1990s, and relatively few fundamentally new pieces of work can be found in recent years. A useful survey is given by [17]. One key question shared by active learning and active perception is how to optimally select new samples in a high dimensional world, a problem that quick becomes computationally intractable. Active learning is also related to the exploration-exploitation dilemma of optimal control and reinforcement learning, i.e. how to trade-off resources allocated to gaining new knowledge vs. resources spent on maximizing pay-off (also called the dual-control problem). A useful recent workshop on active learning in robotics can be found at `http://webdiis.unizar.es/~rmcantin/pmwiki/pmwiki.php/RSS10/RSS10`.

5.1.3 Sensor Fusion and Cross-Modal Integration

One of the **big** differences between biological and artificial systems is that biological systems have massively parallel sensing using various sensory modalities, e.g., haptics, vision, audition, stretch sensors, gyros (e.g., [2]). In most robotics approaches, we overly rely on one or few sensory modality, well, often just because no other one is available. But robustness of information processing most likely arises from sensor fusion using as many sensors as possible, potentially having a huge amount of redundancy. In the vision community, just by intelligent combination of multiple vision processing streams, like intensity, edges, color, etc., one can significantly improve performance [9]. Of course, aligning multiple sensory modalities, fusion of modalities, and weighting them appropriately is a complex topic to be explored in more generality (e.g., [6]). But for the future, instead of relying on rather few modalities and few sensors, understanding how to process many sensory modalities, with numerous and high redundant sensors, how to keep them aligned an calibrated, and how to do all this processing autonomously seems to be a critically missing topic in robotics.

5.1.4 Associative Memories

The human brain seems to be a huge pattern association machine (slightly extrapolating the study of [19]), i.e. sensations in one modality automatically predict sensations or capabilities in other modalities. Thus, anything we do automatically associates what we should perceive, and anything we perceive associates what we can do. The concept of affordances [10] in psychology has this flavor, and research on mirror neurons in the brain [4] plays along the same line of thought. If different sensory modalities can be associated, they could potentially be fused into a more robust percept, i.e. this topic connects also to the previous paragraph on sensor fusion. Building task specific associative memories that span perception and control, that can work on a continuous time scale and also a discrete (more conceptual) time scale, that can be used for inference, planning, and control, would be very interesting. Some related work is in [18, 15] and maybe components of deep learning [5] could be of interest for this topic.

5.1.5 Autonomous bootstrapping of perception and control

Assuming that representations and algorithms for autonomous perception-action-learning systems are somewhat understood, it will be interesting to find methods how the system can automatically bootstrap competence in perception and control over a lifetime. Imitation learning is currently a favored approach to endow a system with some initial biases for motor skills. Bootstrapping perception in an autonomous way maybe more complex as it is harder to provide feedback about the quality of perception. Shaping performance based on some generic optimization criteria would be desirable, although the creation of general optimization criteria is usually quite hard, and the creating of very specific cost functions quickly degrades to “cost function hacking”.

5.1.6 Self-Organization

A fundamental question for perception-action-learning systems concerns their design principles. Is there really hope that a well-managed software engineering approach can do the job, or is it doomed to become a monster like Mircosoft Windows? While a “reboot” in an operating system can be tolerated to some extent, an unexpected reboot in a robotic system could be catastrophic. One could proceed with the realization that radically new design principles will have to be developed for autonomous perception-action systems. Biological systems give us the appeal that mechanisms of evolution, adaptation, goal-oriented self-organization and learning are the key structures. Thus, we would need to focus on algorithmic realizations that can grow automatically, detect structure by themselves, have appropriate reward systems, can self-repair, etc. But it is hard to name successful realizations of this vision. DNA computing might be among the most salient [14], but it is a domain of self-assembly that seems to be largely confined to a DNA world, i.e. DNA systems can only interact with a DNA environment. How this world generalizes to more general domains, even “just” proteins, is completely unclear.

References

[1] Blake A. and Yuille AL. *Active vision*. 1992.

[2] Spence C. and Driver J. *Crossmodal space and crossmodal attention*. 2004.

[3] Golovin D. and Krause A. *Adaptive submodularity: Theory and applications in active learning and stochastic optimization*. 2011.

[4] Rizzolatti G. and Arbib MA. “Language within our grasp”. In: *Trends Neurosci* 21 (1998), pp. 188–194.

[5] Hinton GE., Osindero S., and Teh YW. “A fast learning algorithm for deep belief nets”. In: *NEURAL COMPUTATION* 18 (2006), pp. 1527–1554.

[6] Durrant-Whyte H. and Henderson TC. *Multisensor data fusion*. 2008.

[7] Saal HP., Ting J., and Vijayakumar S. “Active sequential learning with tactile feedback”. In: *International Conference on Artificial Intelligence and Statistics (AISTATS 2010)*. 2010.

[8] Kenney J., Buckley T., and Brock O. “Interactive segmentation for manipulation in unstructured environments”. In: *Robotics and Automation. ICRA ’09. IEEE International Conference on*. 2009, pp. 1377–1382.

[9] Triesch J. and von der Malsburg C. “Democratic integration: self-organized integration of adaptive cues”. In: *Neural Comput* 13 (2001), pp. 2049–2074.

[10] Gibson JJ. *The ecological approach to visual perception*. 1979.

[11] Tsotsos JK. *A computational perspective on visual attention*. 2011.

[12] Itti L. “Real-time high-performance attention focusing in outdoors color video streams”. In: *Proceedings fo the SPIE Humand Vision and Electronic Imaging VIII (HVEI 02)*. 2002, pp. 235–243.

[13] Itti L. and Koch C. “Computational modelling of visual attention”. In: *Nat Rev Neurosci* 2 (2001), pp. 194–203.

[14] Amos M. *Theoretical and experimental DNA computation*. 2005.

[15] Pastor P. et al. “Skill learning and task outcome prediction for manipulation”. In: *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. 2011.

[16] He R., Brunskill E., and Roy N. “Efficient planning under uncertainty with macro-actions”. In: *Journal of Artificial Intelligence Research* 40 (2011).

[17] Burr S. *Active learning literature survey*. 2009.

[18] Lang T. and Toussaint M. “Planning with noisy probabilistic relational rules”. In: *Journal of Artificial Intelligence Research* 39 (2010), pp. 1–49.

[19] Mitchell TM. et al. “Predicting human brain activity associated with the meanings of nouns”. In: *Science* 320 (2008), pp. 1191–1195.

6 The Singularity

Andrew Davison | *Imperial College London*

6.1 Introduction

The aim of this presentation was to look even farther forward than in most of the other sessions, and test the Summit attendees' views with regard to the "Technological Singularity" idea, recently much discussed by "futurists", technologists, science fiction writers and scientists but still on the fringes of mainstream scientific debate. I started the presentation by asking for a show of hands on the following questions:

1. Who knows what The Singularity is?
2. Who thinks it might happen during our lifetimes?

Probably around half of the attendees had at least heard of the idea; and at the start of the discussion I don't think any hands went up in answer to "who thinks it might happen during our lifetimes?". I then explained the Singularity concept, briefly presented the main arguments for it, and we had some discussion. At the end I revisited the second question, and there were three people (including me) out of about eighteen agreeing that it "might" happen during our lifetimes.

A question I had wanted to get to by the end of the session was given that it might happen, what should we do about it? Should foresight of this possibility change the way that we do robotics research? However it was not appropriate to move onto this question when the majority of those present were very sceptical about the whole concept. I must admit that overall I was surprised that such a big concept seemed to be much less on the radar of the world's top roboticists than I might have expected.

6.2 The Singularity

The Singularity is usually defined as an event that could happen in the future when the accelerating progress of technology becomes so rapid that human life is irrevocably changed (for the better or worse). Perhaps the most significant event associated with the Singularity, and certainly the one most relevant to roboticists, is the arrival of human-level artificial intelligence, and then very soon afterwards vastly super-human AI. It seems clear to me at least than human life would not continue in anything like the current way once super-human AI exists. A super-human AI would be able to develop concepts and technologies we cannot understand; would continually self-improve; and for me by definition it would be impossible to control. Various futurists have predicted what might happen to humanity if this were to happen. Some of the possible scenarios are catastrophic of course. The most optimistic scenarios, predicted and hoped for by Kurzweil and others, forsee humans gradually merging completely with their technology, on a path where there might first be great advances in medicine with aspects like longevity expansion, body repair by nanorobots, and brain-computer interfaces. The final destination of this predicted path is “uploading”, where expanded human minds are finally transferred from biological brains to another computing substrate.

This, of course, is far out thinking . . . and as a normal cynical scientist I can understand reluctance to take it seriously. But having thought quite hard about this idea for several years I have not yet found a strong counter-argument which persuades me that it is not possible that something like this will happen in the next few decades. In many ways I would like to hear one! My belief in the idea that the rate of change in technology is accelerating is informed not just by the views of futurists, but strongly by my observation of what is happening in my own field of real-time computer vision, where the new computing power, algorithms and devices we can take advantage of are year on year leading to increasingly staggering capabilities. If we might really be heading towards a Singularity, foresight of this should surely shape everything about the way we are doing robotics research. Let us look at the arguments.

6.2.1 Accelerating Change

The main argument put forward in favour of the Singularity is that the progress of technology has historically followed a law of accelerating change, well described by an exponential curve as a function of time, and that his shows no sign of slowing down. The obvious modern example of this is “Moore’s Law”, originating in a specific observation over 40 years ago about transistor density by Intel founder Gordon Moore, but commonly used to describe the continuing “doubling every 18 months” exponential performance improvement in computer processors and related technology. It is easy to forget how staggering it is that we have moved comfortably, even in my own experience, through talking about computers in the “kilo”, “mega”, “giga” and now “tera” eras.

The idea of the Singularity relies on similar laws describing not just computing performance, but the general progress of technology. The key reason for the rate of progress to keep increasing is that each new generation of technology has the benefit of the last, best so far generation which can be used to develop it. Different areas of science and the

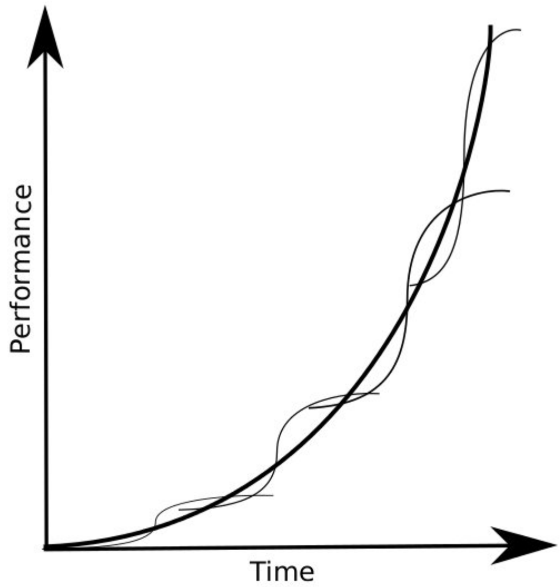

Figure 6.1: A sequence of overlapping S-curved paradigms, each with steeper progress than the last, leading to overall exponential progress.

whole technological economy continually feed back on each other so that every advance makes progress easier in other fields. We are familiar for instance with the idea that scientific supercomputing has revolutionalised many areas of science and engineering through advanced simulation; or that the internet, online research resources and code repositories make it much easier to learn or put together projects in previously inaccessible domains; advances in those areas in turn feed back into the better concepts, designs, methods, factories and robots needed to make the next generation of processors. While some areas of technology do not currently seem to be experiencing exponential progress (e.g., transport speeds?), it can be argued that as more and more industries become information technology and software-dominated they can experience the full benefit of accelerating change – for instance this might happen in manufacturing if 3D printing or nano-technology based self-assembly come to full fruition and an object is just another software file.

Accelerating change is not always smooth, but has been described as taking place via a set of overlapping "S-curves", each corresponding to a paradigm shift (Figure 6.1). A new technology appears; takes some time to break through and overtake the previous model, but then goes through rapid take-up, investment and improvement before eventually reaching the end of its usefulness as some physical limit is approached and slows down; at this point the next technology waiting in the wings, and probably inspired and motivated by observation and desire to beat the previous one, takes over. Even within the narrow domain of digital computer processors, we can see this happening; the increases in performance of single-core CPUs are slowing down, but massively parallel processors such as GPUs, now easily programmable with tools like CUDA, have taken over the exponential progress in terms of the most processing capacity obtainable for commodity prices and are leaving CPUs far behind.

Singularity advocates argue that paradigm shifts have been occuring at an accelerating rate not just in the recent past but right back through human history and beyond into evolutional times. Consider the follow sequence of advances: origin of life, cells, reptiles, primates, upright primates, homo-sapiens, art, agriculture, city-states, writing, printing, industry, electricity, computer, internet, smartphone etc. The "time to next event" vs. "time" plot of these events on a log/log scale is approximately linear. Could we argue that each of these paradigm shifts has similar importance, and just that culture and technology have overtaken biological evolution as the main agents for change in the world (because they happen so much faster)?

6.2.2 Super-Human AI

If we accept the "Strong AI Hypothesis", that general intelligence of the type humans display is achieved by an algorithmic process which can in principle be simulated by an artificial digital processor of sufficient speed and memory, then a continuing exponential increase in computer technology implies that we will at some point all have computers on our desks (and not many years later in our pockets) which are as powerful in raw processing and storage terms as the human brain. Rejecting the Strong AI Hypothesis implies either a spiritual/mystical view that the human mind is more then the brain, or a Penrosian type belief that deep in the brain there might be important processes which are not described by current physics. I think that most modern roboticists are inclined to accept that the brain can in principle be simulated by a Turing machine-like digital computer.

So if we do accept that, then the question of course is how powerful would that computer have to be that could simulate everything a human brain does with enough fidelity to produce the same kind of "generally intelligent" behaviour, and how soon will we have it? I guess that depends on how fiddly you think things are, and how closely you would have to model what each element of the brain does – at the molecular or chemical level, or just at the gross level of unit connections and dynamics. I don't think anyone knows the answer to that precisely yet, but my guess is the latter, that it is the general pattern of connections and signal types that is important. And so if you believe that you have to figure out what level of processing you'd need to simulate that. I've heard recent estimates like

10^{15} calculations per second. Interestingly in his 1950 paper "Computing Machinery and Intelligence" at the dawn of AI, Turing considered that memory capacity, rather than processing speed, was the main limiting factor for AI, and that estimated that 10^{15} bits would be the memory capacity needed to model a human brain. These figures of 10^{15}, whether for processing operations per second or storage capacity, must have seemed stupendous to Turing but for us in 2011 these are "around the corner" figures for desktop computing, coming in 10-20 years surely without much doubt by projecting current curves. But a very significant point is that even if these figures are very wrong, like a million times off, then Moore's Law means this this only makes a difference of a few years. I agree that if it's more like 10^{100} that you'd need to simulate at a much finer level then we are still miles away!

I don't think that even the idea that we will fairly soon have computers available with equivalent computational capacity to the human brain was controversial at the summit. The doubts were naturally about whether we will have any suitable software to run on these devices to achieve human-like "Artificial General Intelligence" (AGI). Where is this fiendishly complicated software going to come from? It might come from a continuation of current AI research and lots of components of the type I develop in computer vision for instance, all joined up with machine learning, some kind of embodiment and training environment, and that something amazing and open-ended happens when you reach the right scale. We have already seen with phenomena like Wikipedia how quickly a vast amount of knowledge can quickly be assembled by relatively uncoordinated means and made easily accessible. Of course this information is not yet in AI-understandable form, though there are efforts in this direction. But we see how current communications and storage capacity might form the way for a set of self-learning and communicating AI agents/robots to build up and share a vast store of "general knowledge". The time, probably near in the future, when commonly available computers achieve the capacity to "in principle" simulate the human brain is a particularly important event in my opinion. Although we still see many weaknesses in current AI and robotics systems, we have never yet run those systems on computers with the processing and memory capacity that we estimate the human brain has; but we will be able to do so soon. Some "magic", emergent, behaviour may happen with the current techniques we have for machine learning, vision, etc. when applied at this scale; or maybe not. It may be that AI still needs serious re-invention, and that we are missing vital algorthmic pieces in the software which might be needed for very general reasoning. But surely we should not dismiss candidate algorithms for AGI (Artificial General Intelligence) until we have tested them at this scale.

Another completely different approach to the software problem might be much more like reverse-engineered biology, where continual improvements in brain scanning give us eventually the ability to see how a whole brain is wired together at the neuronal level and therefore to simulate it functionally on a computer of sufficient capacity. This architecture could then possibly be reverse engineered, improved and expanded.

An important point is that even if your software is not very good, perhaps vastly inefficient compared to what the brain is doing, exponential growth in computing might mean that this just means the need to wait a few more years for the right processor to run that software on. Another point in favour of the singularity that I find convincing is that the human brain, with its capabilities which are dramatically more advanced than those of any other animal, evolved really extremely rapidly to separate us from our common ancestors with the other apes. Surely whatever the human brain does is special is a one particular evolutionary trick (extra piece of code) just massively scaled up and repeated. Personally I am with others in thinking that what we can do that other animals can't is all about a massive, well-organised memory store which we are able to use constantly for prediction.

6.3 Reactions and Discussion

This is clearly a controversial topic, and as I said in the Introduction, after explaining and discussing these ideas with the group, I took a final show of hands where there were three of us with the opinion that the Singularity "might" happen within our lifetimes. To repeat, natural cynic though I am, I find it hard to find any strong argument that the concept doesn't at least merit very serious thought.

There is a tendency I think for scientists working deep in a discipline to see all the local difficulties of that area and to extrapolate those into the future in a kind of "scientific pessimism". For instance, much of the discussion in this summit was about robot manipulation problems, with the perception and planning challenges they present, which is where I think we were mostly agreed that the main thrust of exciting robotics research will be focused over the next few years. There were views expressed that the very high dimensionality of the reasoning needed for manipulation or learning about manipulation makes it doubtful about whether we are making any progress at all on these problems at the moment, and also doubtful whether we would be likely to in the "seemingly short" next 20-30 years.

But history teaches us that our "intuitive linear" view of progress is usually overtaken by historical exponential growth. When a new technology appears we often overestimate the effect that is will have in the near future; but then dramatically underestimate the long-term effect. If half-way through the time allotted to an imformation processing project we have only completed 1% of the planned goals, this might actually be right on schedule for full completion under an accelerating change regime (the human genome sequencing project was an example of one about which there was much pessimisim based on early slow progress, but then was suddenly and surprisingly completed ahead of schedule).

Let's remember that we have only had computers at all for 60-70 years; and only had them at home for around 30; and look at how much has been achieved. People say that AI hasn't come anywhere, but AI is just the moving target of things computers can't yet do; there are plenty of problems that computers couldn't do once where it seemed like "real intelligence" was needed (Being a travel agent? Vacuuming a floor? Route planning and even autonomous driving? Face recognition?) which are now just considered computation. Of course AI has achieved a lot already.

7 Effects of the Funding Environment[1]

Cécile Huet | *European Commission – Cognitive Systems, Interaction and Robotics*

The intention of this talk was to present the activities we are funding in the area of cognitive systems and robotics in the EU research programme and to take the opportunity of having the top level researchers in the field to discuss with them ideas to better serve the community. This was indeed a unique opportunity to learn from these key figures, visionary and experienced, representing the international scene in Robotics. The level of the panel was impressive. Having a panel of international researchers added the global perspective and allowed to benefit from the experience carried out in the US for instance. We acknowledge that the panel represented more a scientific perspective, given that its members were all academics; however we also acknowledge that given their experience, they have a good grasp of the industrial landscape and needs, through cooperation with industry, launching of spin-off companies, etc.

Several ideas have been discussed and the main outcomes are summarized below:

7.1 First outcome

In the coming call for proposals, one of the targets is: "speeding up progress towards smarter robots through targeted competition" The goal is to use such competitions as a tool to support sciences, with a view of providing a mechanism to objectively compare results and measure progress as well as sharing results. To avoid that such initiative becomes an engineering exercise, during which participants tune their algorithms and systems to win the competition, rather than focusing on demonstrating scientific progress.

It is important to run the competition several times and to make sure over-fitting on the data-sets is avoided; the definition of metrics is also key. Tuning the level of difficulty is also critical: it has to be hard enough to connect to deep issues, but not too hard to be relevant to current methodologies. The example of the challenges run by the PASCAL project[2] was discussed. The intention of that project was to “exploit the competitive nature of many researchers and drive progress forward at a faster pace. The challenges could be organized by industrial and academic institutes and would promote the creation of a new generation of more effective methods, demonstrated on real-world problems. The results are published as papers and presented at PASCAL sponsored workshops, in

[1] Disclaimer: The information in this document does not necessarily represent the official view of the European Commission (EC).

[2] `http://www.pascal-network.org/?q=node/15`

addition to some of the major conferences in the field." This was considered as a breakthrough and changed the field when introduced, but now it seems that it has reached a stage of minimal improvements through tuning. So it is important to learn the lessons from that example. Such challenges or competitions have the potential to increase the visibility of the field and also give the opportunity to small players to be highlighted. In conclusion, this is a delicate initiative but could play an important role in the field, for the various stakeholders. Therefore we have to be careful in guiding the proposers, during the selection and the negotiation of the potential proposals in order to best exploit such potential.

7.2 Second outcome

Through our activities, we also try to stimulate the cooperation between the academia and the industry, in particular supporting the transfer of knowledge and finding mechanisms generating win-win situations. The intention is to initiate a snowball effect to gradually develop the exchanges between the research results and the industry needs.

To that end, through our next call for proposal we expect to fund projects aiming at "gearing up and accelerating cross-fertilization between academic and industrial robotics research", through definition of joint industrially-relevant scenarios, shared research infrastructure, experimentation with industrial platforms, benchmarking activities, etc. A possible mechanism to reach that goal could follow the example of the existing project ECHORD[3].

It was interesting to hear the perspective from the academic side regarding such activities, and in particular ECHORD. The feeling is that such initiative could be good. From the experience of the panel, in general academia-driven cooperation tends to be more successful than industry driven, so is the cooperation with SME as opposed to cooperation with large industry as experienced in the US, in the context of earlier programmes encouraging academia-industry cooperation. There were also some concerns raised regarding the current model of ECHORD, some expressed concerns regarding the selection and monitoring process which should ideally be run by an independent body, such as the EC, as opposed to research laboratories, also due to the infrastructure that has to be built in order to run such projects. Another concern, from the academic perspective, is the scientific contribution of such experiments.

The community can be reassured regarding the fairness and un-biased selection process, in which the EC is involved and carefully monitors all the steps, which are in line with the rules applied for EC selection process. Regarding the scientific contributions,

[3] `http://www.echord.info/`

this will have to be demonstrated by the running experiments. Lessons will have to be learned from the first set of experiments in order to optimize both industrial and scientific benefits. This is a key element to keep the level of involvement and commitment form the academic side.

7.3 Third outcome

This event was also timely to expose to the experts the future EU funding programme for research and innovation, covering the period 2014-2020: "Horizon2020". We are indeed in a consultation process to prepare this future programme, so the timing is ideal to collect ideas about topics and potential new mechanisms to support the field, including areas of international cooperation.

Regarding possible international cooperation in basic research, the intention is to build on excellence available in the world for more efficient scientific progress. However, it was stressed that we should not try to avoid duplication in basic research, since competing approaches breed new ideas. Regarding areas of cooperation, we should not be too directive, but rather let cooperation emerge. In terms of standardization, it was felt that at this stage, it is more important to try to provide common tools rather than trying to establish standards. Mechanisms to foster international cooperation were also discussed and exchanges of researchers at all levels (e.g., Max Planck Guest Professorship programme[4]) as well as summer schools were considered as powerful tools. It is however important to find simple and flexible procedures to implement these.

The concept of a shared infrastructure, possibly in form of a network of "Physical Nodes – Robotics Living Labs" was also discussed as another potential tool to support the community. The intention would be to give access to researchers to robotics platforms and maintain them, to offer testing and validation infrastructure, which could also be used as a means to disseminate results and have show-cases to demonstrate to potential users and industry the capabilities of today's technologies and research results. Later on, this could also serve as certification center, for robotics-related products or services. However several concerns were raised regarding this idea in particular the overhead to run such facilities and the doubts regarding the possibility to develop the software remotely from the hardware (either using remote access to the hardware or testing with simulators and having occasional access to the hardware).

Several provocative questions were also raised to trigger discussions in view of identifying how to better serve the community (both the research community and the European

[4] http://www.research-in-germany.de/info/senior-researchers/funding-for-senior-researchers/funding-programmes-senior-researchers/58820/max-planck-foreign-visiting-scientists-at-max-planck-institutes.html

Industry)

How to make multidisciplinary international cooperation work – or are forced marriages wrong? How to demonstrate that the field is progressing? How to make sciences meeting the industry and industry helping scientific progress? How to create a snowball effect between the offer from sciences and the needs from industry? How is the academy-industry couple doing? More generally how to transfer scientific results to real life? Why is there little progress in service robotics despite all the research efforts?

Artificially forcing people to work together is not the right approach. One of the reasons why there is little progress is because hard problems have to be solved to do something useful in service robotics (see for instance limited dexterous manipulation capabilities of current robots, in particular when it comes to daily tasks for non-fully specified tasks and environments). Selected projects tend to focus too much on showing capabilities (leading to hacks) rather than exploring new, basic ideas that will take time to show impressive results for the wider public. Funding mechanisms should be more open to such exploratory activities.

It was considered a good idea to target as interim goal, for the field of robotics in general, the deployment of service robots in real production (e.g., surgery, maintenance, logistics, SME manufacturing, dangerous/heavy tasks) as a step towards the longer term vision of having pervasive useful robots, exploiting robotics technologies to solve "time wasting" problems and to supporting the ageing society.

Several open questions were finally raised not all of them were addressed due to time constraint: How to better link/align the EU Programme to Education? Could multidisciplinary curricula help? How to better link/build synergies between EU-National programmes? How to build synergies – optimize resources? How to reconcile the requirements of EU projects versus Academic career and versus PhD training? How to assess progress Academic/research/industry/users?

How to foster excellence through European projects? How to make it more competitive during projects? Instead of the suggested go/no go decisions mid-term in the projects, depending on the achievement of their targets, it is suggested to implement a ramp in support depending on the success.

In conclusion, the questions raised triggered very interesting discussions and ideas on concrete mechanisms to better serve the community. I have reported on them and we will take them into account, to the extent possible as input to our consultation process.

8 Plastic Maps for Life Long Navigation

Paul Newman and Winston Churchill | *University of Oxford*

8.1 Thoughts On Arriving at the Summit

I don't know how to do this, but I know we must - our machines must be plastic[1]. Not physically of course, but in terms of perception, scene understanding and performance. We want our robots to get better through everyday use. Sure we can demand that we manufacture them with certain baseline competencies that fulfil worse case operational requirements but that is in no way sating. Say I had a car which could drive itself some of the time, so I spend my mediocre salary on buying such an robot so I can have part of my day back. If I drive that car every day on the same commute I expect it to be able to do more and more of it autonomously each day. Generalising away from autonomous transport, I want machines to learn on the job, I want it to stretch base line performance by executing and analysing a core competency again and again, it should mould itself to its daily surroundings. And finally to add one more simile to plastic it should be durable. This for me is lifelong learning. The tricky part is finding a problem domain that is narrow enough to avoid ultimately equating lifelong learning with "thinking machine" so one can get started. Vast scale, vast time navigation seems plausible but could anything learnt here be generalised? Will it all end up being a mixed bag of tricks?

The above is what I wrote before the meeting and I think I got the emphasis wrong – figuring this out was a great outcome of the meeting. You see, in my mind I equate longevity with plasticity. One could only have a machine navigate for years if it maintained a plastic representation of the world - one that moulded itself to new data and could accommodate slow changes of appearance and structure. Now, the solution I had in mind to achieve this is one of accumulation of experience and as I explain below this will work (indeed on leaving the summit I wrote a paper on just that). However it became clear to me that these ideas do not easily migrate to other domains - especially those like manipulation and planning. I left wondering if the navigation problem was in some way special. Certainly I was not so convinced that simply "recording everything" is going to help the manipulator guys or the planning guys. But I was even more sure I wanted to try it out for large scale navigation . . .

[1] `http://www.youtube.com/watch?v=CsrLHP26zvk`

8.2 Thoughts On Leaving the Summit

Without doubt the summit shaped my thoughts and without doubt it was some of the most valuable days of research/talking time I have had in the past few years. The format was extraordinary – we had time to engage, we had an excellent group of people who knew all the right things and the venue was ideal for this kind of scholarship. I left with a clearer picture of what I was actually advocating and set about writing a paper that made that crystal clear and demonstrated it working. I fear that without the summit the point I wanted to make would be adorned with miscellaneous distraction. So here now, on the back of that summit is where I stand on this issue of plastic maps now:

To achieve long term autonomy robotic systems must be able to function in changing environments – we see this as a big challenge. Change can come from many sources: sudden structural change, lighting conditions, time of day, weather and seasonal change. To illustrate, consider the problem of ego-motion estimation with a camera mounted on a robot operating outdoors. This is a richly mined area of research and immediately we reach for a visual navigation (SLAM) system that can map and localise all at once. But what should we do if we revisit a place and its appearance has changed drastically – perhaps it has snowed? What do we do if a place's appearance slowly creeps as summer turns to autumn? Should we undertake some unifying data fusion activity to yield a monolithic map in which we can localise? We argue that we should not; in the limit such a map would have to contain features from every possible scene modality. The things we see on a given tree in winter are simply not the things we see in summer; the details we see on a wet road at high noon are different to those we see at dawn when the road is dry. We shall not force things to be coherent. If, for example, part of a workspace on Tuesday looks wildly different on Wednesday then we shall treat these as two independent experiences which equally capture the essence of the workspace. We shall only ever tie them together topologically.

A high level view of our approach is appropriate here. On the initial visit to a new area we save a constellation of visual features like most systems. For reasons that will become clear we call this an "experience" rather than a map. When revisiting the area the robot attempts to use the live stream of images to localise in the saved experience. If at any point this is unsuccessful, a new experience is created based on the current appearance of the world. As the robot continues, still saving to this new experience, it is also trying to re-localise in its previous experience(s). If this is successful at any point, saving is stopped and the system returns to localising in its previous experience. Importantly this methodology causes the system to "remember" more representations for regions that change often, and fewer for regions that are more staid. We call the collection of all experiences the plastic map. Note that we handle new routes and complete localisation failures seamlessly – indeed it is the failure of localisation which drives the saving of a new experience. This is because we make the assumption that our localisation fails because of bad or unsolvable data association – what was there before is simply not there now.

A core competency on which we depend is a visual odometry (VO) system which continuously produces a (possiblyephemeral) 3D model of the world using a stereo pair. This system is always on, always consuming the live stereo pair stream and estimating the relative transformations between camera poses and producing 3D feature locations relative to camera poses. Concretely an experience is a stored set of relative poses and feature locations. Note the emphasis on relative; we entirely avoid operating in a single global frame. All we require is an ability to render a metrically correct idea of camera motion and 3D feature locations in the vicinity of the robot's current pose – we do not care about the location of things that are far away and which we cannot see. Upon revisiting an area, localisation is attempted in all previous experiences that are relevant to the area.

By keeping experiences independent we are able to run a "localiser" for each. This can trivially be done in parallel and allows the system to utilise relevant experiences. In reality, at runtime we see that the number of active and successfully localised experiences is small. After all, each new experience is only created out of necessity because it is visually different from all others. Therefore subsequent visits to an area should be able to localise in only a small number of experiences as they are by construction visually different. Finally we would like to stress that although we describe the framework using vision, it is actually agnostic to the sensing modality and could be used with other sensors such as laser range finders so long as equivalent systems to the ones described above are supplied.

We have tested our system on 53 runs of two laps of a 0.7 km circuit, covering 37 km in total and consisting of over 136000 stereo frames. The data were collected over a three month period at many different times of day and in different weather conditions.

At the moment it does seem to be working . . .

9 Towards High-Performance 24/7 Cognitive Humanoids

Tamim Asfour | *Karlsruhe Institute of Technology*

Recently, considerable progress has been made towards the realization of humanoid robot systems which are able to move in a human-like way and perform tasks in human-centered environment. However, current systems are still limited in their actuation, sensing, prediction, interaction and learning capabilities. We define High-Performance Humanoids as integrated complete humanoid robot systems able to act, interact, predict and learn in 24/7 manner in the real world and to perform a wide variety of tasks.

9.1 State of the Art

In recent years there are renewed efforts to develop robot systems that can perceive, move and perform actions. An encouraging spectrum of many isolated elements in the area of cognitive systems has been realized with a focus on performance in well-defined, narrow domains. The development of cognitive robots relies on artificial embodiments having complex and rich perceptual and motor capabilities. This leads to robots with rich sensorial inputs and complex actions necessary to develop higher cognitive processes. These aspects are, thus, particularly supported by humanoid robots ([7, 6, 3, 2, 4, 14, 11, 16, 13, 1, 12, 5, 15, 10, 8]), i.e. (embodied) robots that perceive, move, and perform diverse actions, which are often acquired by learning techniques.

Although current systems are technologically advanced, they are not able to learn in an open-ended way and their behaviours and lifelong learning capabilities are limited. Successful attempts in building complete systems are still limited to systems designed for "sunshine" environments with limited scope and simple tasks in a given scenario. The transferability of the developed skills and abilities to varying contexts and tasks without a costly redesign of specific solutions is still impossible. Complete robot systems integrating perception, action, planning and lifelong learning capabilities, which are necessary to interact with the environment, as well as a variety of functionalities which are needed to carry out diverse tasks in real environments are still missing.

Recent examples of complete systems which are able to perform a variety of tasks are the Willow Garage Personal Robot PR2, a wheel-driven robot with two arms which is able to perceive its environment detect wall outlets and plug itself in for recharging. The Twendy-one robot (see [8]) developed at the Waseda University in Tokyo posses a wide range of capabilities in human environments such as carrying a tray, fetching objects from the refrigerator, etc. Similar tasks have been presented using the HRP-2 humanoid robot series in [12]. The humanoid robot ARMAR-III (see [1]) is endowed with a variety of capabilities to perform tasks in a kitchen environment such as grasping daily objects, loading the dishwasher and fetching objects for the refrigerator as well as learning various behaviours from human observation. On the humanoid robot DB at ATR, Japan, various behaviours have been demonstrated such as paddling a single ball on a racket, learning a folk dance by observing a human perform it, drumming synchronized to sounds the robot hears (karaoke drumming), juggling three balls, performing a Tai Chi exercise in contact with a human, and various oculomotor behaviours [2]. However, although these robots can perform different tasks, they do not possess the ability of autonomous, life-long learning, which is crucial for robots with the ambition to operate in human living spaces. Although learning has been employed to acquire single tasks, the applied learning techniques were specialized and did not consider the problem of lifelong learning from sensorimotor experience.

9.2 Challenges and Research Roadmap

Today, humanoid robots can be considered as highly advanced mechatronics systems with complex and rich sensorimotor capabilities. Thus, such systems are as the most suitable experimental platform for studying human behaviors and cognitive information processing. In the following some of the challenges towards the realization of high-performance humanoids are briefly discussed.

9.2.1 New Bodyware for Humanoids

Research efforts related to the development of high-performance humanoid robots must address the question of how the body morphology must support processes and representations for emergence of cognitive capabilities.

- The design of humanoid robots with human-like capabilities requires a new thinking regarding design mechanisms, materials and control. Novel technologies and methodologies are needed for the development of compliant, high-performance and energy efficient actuators, sensor technologies (in particular skin), soft materials, as well as dynamically reconfigurable software and hardware architectures and high density lightweight power sources.
- Investigation of design principles and quantitative models for the development of

systems that

- explore their own sensorimotor primitives and body morphology
- explore the environments and the effective interaction with it
- predict the body dynamics and the physics of the world

- How body morphology allows to cope with morphological change arising through the interaction with the environment and tolerance to uncertain variability in performance of single robot components.
- How reconfigurability and self-reconfigurablility, redundancy, robustness and flexibility in technical systems can be implemented.

9.2.2 Objects, Actions and Prediction

To deal with problems on perception and action researchers in the late 80s introduced two new frameworks with parallel efforts, in the field of Computer Vision and the field of AI/Robotics, under the headings of Active Vision (Animate, Purposive, Behavioural) and Behaviour-based robotics respectively. In both formalisms, the old idea of conceiving an intelligent system as a set of modules (perception, action, reasoning) passing results to each other, was replaced by the new idea of thinking of the system as a set of behaviours. Behaviours are sequences of perceptual events and actions. These efforts still go on, but have only led to limited success. One reason for that is that such perceptual events involve recognition which is such a hard problem that it prevented the new formalisms from making a breakthrough. A further reason for failure is that the behaviours were never meant to involve objects into the action (e.g. for recognition). A third reason was that no one managed to formulate any theory for behaviour-based robotics. Hence, it was impossible to predict how the systems developed would scale up and deal with new situations.

Human understanding of objects is essentially multi-sensorial. It develops during an intensive exploration making use of visual and haptic information. Therefore, the cross-connection between haptic and vision must be analysed. The gained knowledge would play a key role in modelling multi-sensorial processes in artificial cognitive systems, which then can develop a more holistic understanding of the perception-action coupling and thus objects and actions. In traditional information theory, the environment only plays the part of a passive, undirected disturbance (for example also in closed loop control theory) negatively affecting the input-to-output transfer characteristics of a system. Here we propose, instead that the environment should be treated as an active component. It is active through "my own actions" (the actions of ego) and those of "the others" (the actions of alter), which feed back to ego. Thus, traditional information theory is not sufficient to describe the interaction of an agent with its world correctly. Instead this problem needs to be addressed in a closed loop paradigm where ego acts in its environment and observes the consequences of its actions (in interrelation also with alter). This notion has entered modern robotics theories by the qualitative term "rootedness", which refers to the necessity to embed an artificial acting agent in an environment. Thus, instead of using the conventional I/O paradigm, new approaches should introduce so called encased closed loop situations defined by the mutual interactions of an organism (ego) with its environment. An encased closed loop describes a conventional sensor-motor feedback control loop but with an active environment monitored from the perspective of ego. This represents a central shift of paradigm and follows a constructivist's viewpoint where the environment becomes an integral part of the system's description. This notion goes clearly beyond the conventional concept of a perception-action loop. It embeds the agent into its environment and into its social group by the same formalism. On the side of theory this will lead to intrinsically consistent and technologically applicable measures of "autonomy", "contingency", and "complexity" of agent-world- as well as agent-agent interactions, resulting in the first steps towards an information theory of encased closed loops.

Research into cognitive robots should combine the study of perceptual representations that facilitate motor control, motor representations that support perception, and learning based on actively exploring the environment and interacting with people that provides the constraints between perception and action. This will then allow, e.g., to learn the actions that can be carried out on and with objects when making use of the interplay of different sensorial modalities, such as vision, haptics and acoustics. Action-centred cognition presupposes that artificial cognitive systems will be equipped with eyes, sophisticated haptic sensors for its end-effectors and microphone-ears. This allows for efficient interaction with the world making use of the full potential of multi-sensorial representations.

Object-Action Complexes

The European project PACO-PLUS (Perception, Action and Cognition through Learning of Object-Action Complexes, www.paco-plus.org) has introduced the concept of Object-Action Complexes (OACs) to emphasize the notion that for a cognitive agent objects and actions are inseparably intertwined and that categories are therefore determined (and also limited) by the action an agent can perform and by the attributes of the world it can perceive (see [9]). The resulting OACs (pronounced "oaks" are the entities on which cognition develops (action-centered cognition). Entities "things" in the world of a robot (or human) will only become semantically useful objects through the action that the agent can/will perform on them.

The OAC concept is based on, but extends the Gibsonian concept of "affordance". In contrast to constructivist approaches, Gibson claimed that objects and events in our environment provide an actor all the information they need about the actions they "afford". This claim was motivated by the idea that perception is not a static process, but rather is a temporally extended act of information acquisition. In other words, the affordances our environment provides are revealed by actively exploring it.

PACO-PLUS has made significant use of the Gibsonian approach in two respects. First, PACO-PLUS has made the robot an information-seeker that is actively experimenting with the objects it is facing in order to find out more about its perceptual features and the action opportunities they provide. In other words, the PACO-PLUS agent is no longer a passive knowledge receiver but an *active explorer*. Second, the idea that active information acquisition reveals the objective structure of our environment makes it possible to ground cognitive representations. Once objective environmental information about sensorimotor opportunities is encoded in the lowest level OACs, these OAC level provide a reliable basis for forming higher-level, more abstract representations suitable for reasoning and action planning, while still being grounded in the robot's sensorimotor experience.

9.2.3 Representations

Building humanoid robots able to learn to operate in the real world and to interact and communicate with humans, must model and reflectively reason about their perceptions and actions in order to learn, act, predict and react appropriately. Such capabilities can only be attained through physical interaction with and exploration of the real world and requires the simultaneous consideration of perception and action. Representations built from such interactions are much better adapted to guiding behaviour than human crafted rules and allow situated and embodied systems, such as humanoid robots in human-centered environments, to gradually extend their cognitive horizon. Such representations should allow for learning and extending representation in ways that transform intractable problems into tractable ones and support generalization and knowledge transfer between different cognitive systems. These representations should take into account space and motion, objects (things that move) and actions, properties and affordances, goals, plans, beliefs and desires, communication, and models of other minds. In this context several research questions must be addressed.

- How to extend and improve exploration-based and stimulus-driven knowledge acquisition?
- How to define the actual algorithmic mechanisms by which an agent can generalize knowledge across domains leading to a generative extension of its experience?

- How to embed these two mechanisms in a dynamically stable process to drive the extension of knowledge in a generative way while interacting with its environment and other agents (humans)?
- How to allow the agent to predict its own perception-action loops, but also - importantly - the actions of other agents, leading to advanced abilities to cooperate, interact and communicate?
- How to integrate exploration-based and generative inside-out processes into an advanced, complete embodied cognitive system?

9.3 Examples for Research Challenges

9.3.1 High-Performance 24/7 Humanoid for daily life

The challenges of creating high-performance 24/7 humanoid robots can be summarized as follows

- Understanding and interpretation of scenes, contexts and situations
- Categorization of daily objects
- Grasping and manipulating any object (Pin, book, . . . , beer box)
- Navigation in every environment (Home, street, super market, etc.)
- Human-Robot interaction
  - Multimodal interaction
  - Physical interaction
  - Natural communication
  - Action and activity and intention recognition
  - Human tracking, gesture detection, face detection and identification, emotion recognition
- Social interaction (Humor, trust, privacy)
- Personalization: Adapt to humans needs and habits.

What to measure?

Criterion	
Energy consumption	Similar to other household appliances (oven, fridge, dishwasher, etc.)
Program complexity	FLOPs, Memory requirements
Performance	2015: set/clean the table, load the dish washer/washing machine, prepare food. 2030: Clean the apartment, go shopping (in super market, Italian shop, etc.) 2049: Similar to human caregiver in performance and social interaction
Price	Cheap car

9.3.2 High-performance humanoid robot that can play tennis

It is not about tennis but about the following scientfic and technological challenge:

- Understanding the body dynamics, body balancing and motor coordination
- Safe falling and recovery
- Real-time prediction: reaction based on vision would be too late. *Sense-Plan-Act* would not work. Instead **Predict-Act-Sense**.
- Learning of others behavior and adaptation of own behavior based on past experience and learning to predict and adapt from little experience and few examples.
- Multisensory integration (vision, vestibular, haptics, . . .)
- High speed perception and high speed control.

What to measure?

Criterion	
Energy consumption	Humanoid robot should be able to play a game with the energy equivalent of a ”Maultaschen” dish.
Program complexity	FLOPs, Memory requirements
Performance	2020: Perform basic tennis playing 2030: Steadily win against number 500 of the ATP ranking 2049: Steadily win against number one of the ATP ranking
Price	Cheap car

References

[1] T. Asfour et al. “ARMAR-III: An Integrated Humanoid Platform for Sensory-Motor Control”. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*. Genova, Italy, 2006, pp. 169–175.

[2] C. G. Atkeson et al. “Using Humanoid Robots to Study Human Behavior”. In: *IEEE Journal on Intelligent Systems* 15.4 (2000), pp. 46–55.

[3] R.A. Brooks and et al. “The Cog Project: Building a Humanoid Robot”. In: *The First International Conference on Humanoid Robots and Human friendly Robots*. Tsukuba, Japan, 1998.

[4] G. Cheng, A. Nagakubo, and Y. Kuniyoshi. “Continuous Humanoid Interaction: An Integrated Perspective - Gaining Adaptivity, Redundancy, Flexibility - In One”. In: *IEEE Journal on Robotics and Autonomous Systems* 37.2-3 (2001), pp. 161–183.

[5] G. Cheng et al. “CB: A Humanoid Research Platform for Exploring NeuroScience”. In: *6th IEEE-RAS International Conference on Humanoid Robots*. 2006, pp. 182 –187.

[6] K. Hirai, M. Hirose, and Y. Haikawa nad T. Takenaka. “The Development of Honda Humanoid Robot”. In: *Proceedings of the IEEE International Conference on Robotics and Automation*. Leuven, Belgium, 1998, pp. 1321–1326.

[7] K. Hirai et al. “The Development of Honda Humanoid Robot”. In: *IEEE International Confernce on Robotics and Automation*. Leuven, Belgium, 1998, pp. 1321–1326.

[8] H. Iwata and S. Sugano. “Design of human symbiotic robot TWENDY-ONE”. In: *IEEE International Conference on Robotics and Automation*. 2009, pp. 580 –586.

[9] N. Krüger et al. “Object-Action Complexes: Grounded abstractions of sensory-motor processes.” In: *Robotics and Autonomous Systems* 59.10 (2011), pp. 740–757.

[10] G. Metta et al. “The iCub humanoid robot: an open platform for research in embodied cognition”. In: *PerMIS: Performance Metrics for Intelligent Systems Workshop*. 2008.

[11] T. Miyashita and H. Ishiguro. “Natural Behavior Generation for Humanoid Robots”. In: *I. J. Humanoid Robotics* 1.4 (2004), pp. 637–649.

[12] K. Okada et al. “Vision based behavior verification system of humanoid robot for daily environment tasks”. In: *Humanoid Robots, 2006 6th IEEE-RAS International Conference on*. 2006, pp. 7 –12.

[13] I.W. Park et al. “Mechanical design of humanoid robot platform KHR-3 (KAIST Humanoid Robot 3: HUBO)”. In: *Humanoid Robots, 2005 5th IEEE-RAS International Conference on*. 2005, pp. 321 –326.

[14] Y. Sakagami et al. “The intelligent ASIMO: system overview and integration”. In: *IEEE/RSJ International Conference on Intelligent Robots and System*. Vol. 3. 2002, pp. 2478–2483.

[15] G. Sandini, G. Metta, and D. Vernon. “50 years of artificial intelligence”. In: ed. by Max Lungarella et al. Berlin, Heidelberg: Springer-Verlag, 2007. Chap. The iCub cognitive humanoid robot: an open-system research platform for enactive cognition, pp. 358–369.

[16] K. Yokoi et al. “Experimental Study of Humanoid Robot HRP-1S”. In: *I. J. Robotic Res.* 23.4-5 (2004), pp. 351–362.

10 Is Robotics the new Physics?

Gregory Dudek | *McGill University*

This is a position statement from the author, meant to generate some debate at the recent Berlin Summit on Robotics. It relates to the role robotics can play as a unifying banner for several areas of inquiry.

The key argument is the potentially primal role that robotics seems destined to play in the intellectual lives of people in the next century. This can be framed in the context of the history of science. Looking back to Ancient Greece, Physics (or "Natural Science" as it was more generally framed) played a critical unifying role in shaping humanity's understanding of the natural world and our role within it. This unified vision was significant in that it allowed for an organized and coherent development of new ideas and source of knowledge (and power) that the Western world went on to develop over the subsequent two millennia; a process of intellectual growth akin to, and related to, developments elsewhere in the world. It has served as the intellectual framework that related mathematics, the physical world and our conception of how it works and how to control it. As Natural Science became known as "physics" (which was essentially synonymous with "all science"), and which eventually bifurcated into the allied specialties that make up the various natural sciences (such as Chemistry and Biology) it allowed scientists to determine what common bodies of basic knowledge constituted their domains, and to develop not only theories and models, but also shared curricula, academic programs and even political agenda.

Our field is now facing the same challenges in the domain of "synthetic science": the science of the artificial, virtual and man-made systems that are already of enormous influence and importance. The constructs of synthetic science (such as a major computer operating system) are already as complex as anything that mankind has ever built, both in purely intellectual terms as well as in terms of actual artifacts. We need to not only develop tools and methodologies, but also to identify and formalize basic questions, and to circumscribe coherent new domains of discourse. Thus, as synthetic science progresses, new challenges are developing based not only on ambitious new goals we want to achieve, but due to the complexity of the objects and ideas under consideration. Much of this new domain of science and engineering can be described by one broad term: robotics.

Robotics in its broadest form can be defined as the discipline concerned with both the development and modeling of systems that (1) make measurements of the real world, (2) perform computations, and then (3) act upon the real world in some substantial way. By this definition, more and more of the objects in our everyday world are becoming robots, and this is happening rapidly. This includes, of course, cell phones, cars, security systems, and many of the appliances in our homes. The microwave oven in my own home, for example, measures the weight and humidity of food we put into it, computes the appropriate cooking time and power levels needed, and then acts upon the food to cook it. As almost every object within our lives becomes computationally enabled, myriad new challenges are starting to emerge. As many devices start to become independently mobile, or interact with other devices that are mobile, these inherent challenges will increase substantially. As our culture is subsumed by robotics technologies, do we not need an all- embracing domain for this huge new body of challenges?

The implications of defining robotics as a broad umbrella are twofold: one pragmatic and one conceptual. The conceptual implications relate to the development and organization knowledge, the construction of pedagogical systems and programs of instruction, and the development of formal mathematical frameworks for very complex artificial or emergent systems. The pragmatic implications relate to the fight for funding, recognition of programs within our universities, and the ability to efficiently carry out our research.

Physics and natural science has been defined as the understanding of the “laws and phenomena of the natural world,” while traditional engineering deals with the application of that understanding to the creation of new artifacts. Our challenge in robotics is also to understand and predict the operative laws in our discipline, but they are not exclusively the laws of the “natural world”, and in fact we have the option to generate new laws (for example network protocols that govern information flow or connectivity). Thus, robotics is profoundly theoretical as well as distinctly experimental.

Moreover, a critical part of the robotics research enterprise is to build, measure and eventually control the artifacts we are envisioning. These steps are not always sequential: with networked systems, for example, we often observe unintended phenomena that must be understood after a system has been designed, built and is already under control.

Not only will (does) robotics impact our conception of the world and our conceptualization of our role in it, robotics also has the potential to impact our very sense of identity. It is a domain that has already impacted notions of how people function and how biological organisms evolve. As such, robotics is reshaping not only our lives and our society in pragmatic terms, but also how we see the world and ourselves within. Is this not the same kind of conceptual reformulation that led to the Renaissance?

What is required is a unifying science of what will govern a critically important new world view. If robotics and related technologies have the impact we expect, and which in fact seems inevitable, then there can be no doubt it will impact our conception of science, engineering and society. One needs only to reflect on how notions of computing, computers and algorithms have shaped most areas of human though over the last 50 years, where computational ideas have fundamentally changed thinking in areas as diverse as biology, banking, dating, sculpture, communications and criminology.

In addition, is it clear that many important ethical and social issues are looming. They need to be addressed in a context that is technically broad and mature.

A topic of current discussion and debate both at the workshop and in society at large is the notion of “The Singularity”, as defined by Ray Kurzweil (see also 6). While the singularity itself is a topic of substantial controversy and some doubt, the accelerating pace of technological change that is used to substantiate this notion is broadly agreed upon. This accelerating rate of change increases the need, and the urgency, of recognizing the role of robotics today and bringing the disparate ideas and disciplines involved into a coherent and collaborative framework.

Robotics is the branch of human endeavor that integrates both engineering and science, and cannot be pegged well in either alone. By subdividing the field into 2 different academic faculties (Science and Engineering) or disparate disciplines (Computer Science, Mechanical Engineering and Electrical Engineering), the additional potential for fruitful interaction is decreased precisely in a subject where this interaction is critical. In addition, it becomes more difficult to recognize a common body of prerequisites, knowledge, and tools that the students and practitioners would best be equipped with. In short, the divide between Science and Engineering is not appropriate to a domain of discourse defined by intellectual constructs that are created by human hands. Robotics, perhaps more than any other area of inquiry, falls on both sides of this divide and thus progress is directly impeded by the partition between traditional engineering and science.

Robotics has a fundamentally different (and broader) mandate from many classical areas of computer science like complexity theory, compilers, quantum information theory.

Much of classical science is reductionistic, but even the scientific part of robotics are not.

Artificial Intelligence, as a research domain without robotics, becomes increasingly arcane and irrelevant. Likewise Computer Vision without robotics would have to ignore fundamental issues of great value and importance.

Traditional academic disciplines like Computer Science, Mechanical Engineering, and Electrical Engineering are likely to be preoccupied primarily with embedded systems, smart machines, and self-diagnostics systems. Systems which are, in a deep sense, robotic systems. Moreover, systems which fundamentally and by their very nature cross the barriers between these narrow disciplines.

No other containment relationship between academic disciplines is as consistent as using robotics to refer to the high-level aggregate. Robotics cannot be a little niche within

Mechanical Engineering or Computer Science, it just does not fit such narrow confines. On the other hand, while no strict hierarchical ordering of academic disciplines is perfect, making Robotics an umbrella discipline for several other sub- areas is probably very natural, and will become more so as the science and technologies of the discipline evolve.

We need to promote this coherent world view in education and government. Such a unified framing of the discipline is useful in the quest funding, student development, the consideration of ethical issues and other integrative-level issues.

11 The Robot Human

Roland Siegwart | *ETH Zürich*

Despite the heavy investments into R&D in robotics and the important progress of the last years, robots are still lacking far behind humans when it comes to scene understanding and complex navigation and manipulation tasks. Even though that the research community promised since years, that robots will soon be able to support us like a butler in our daily life, we are still far from having robots that can reliable operate and fulfill tasks even in simple environments.

However, what science fiction movies already "created" and presented many years ago is gradually becoming reality – humans are converting to cyborgs, thus converting to the most sophisticated robots. This offers novel concepts and opportunities for getting humans' help for service robots or even us humans as service robots – the Robot Human.

Today, humans permanently carry sensor and communication devices with them that allow measuring their actual position (GPS), their motion activates (IMU), their currently perceived environment (camera) and much more. Soon also mental activities and physical / health status will be available. Furthermore, humans are, e.g., through Facebook, permanently on-line, ready to take message or even orders from the world-wide internet community. And incoming information about local offers (e.g., parties, restaurants ...) are easy available through iPhones and other devices. Thus our behavior and activities are more and more controlled through these devices and the internet community. Therefore humans becoming controllable devices like robots that can be used for fulfilling tasks and services. Potential human service providers (e.g., I can guide you through Berlin) and service requesters (e.g., I need a tourist guide for Berlin) can be matched according to their collocation and activities. The usage of the internet community for service tasks has already been implemented for text recognition or image annotation, and tele-operation and tele-presents of humans is gradually being implemented.

Up to recently, robots had the big advantage over humans, that they could easily share knowledge over the internet. However, thanks to recent technological advancements in mobile communication and smartphone technology, humans are today permanently on-line and are also sharing knowledge on a fast pace. A striking example is the organization of political movements, like the recent revolutions in Arabic countries that were "controlled" through Facebook and Twitter. Another typical example is the GeoCaching game, through which a large community is fulfilling treasure-hunting tasks free of charge – just motivated by the fun of it. In both cases, humans can be considered as autonomous agents that are "controlled" very efficiently through a self-organizing network. Considering the fact, that kids, especially boys, spend many hours each day in the cyberspace and on social networks, the potential of “the robot human” is huge.

Robotics technology is of key importance for these developments and already making major contribution to it. It is expected that humans will become important service provides for robots or other humans, e.g., be helping robots in scene understanding and contextual learning. It has recently been demonstrated, that object recognition and image annotation of images perceived by robots can successfully be solved in nearly real-time through posting an inquiry on the internet. In the near future similar approaches might also help robots to select appropriate action in complex setting or learn complex manipulation tasks – just by asking for help through social networks.

There is no doubt about the potential of taping humans competences for solving tough robotics problem or even to use humans as “robot humans” for offering services to society. This development has already gained momentum and will surely shape the future of the human-robot society.