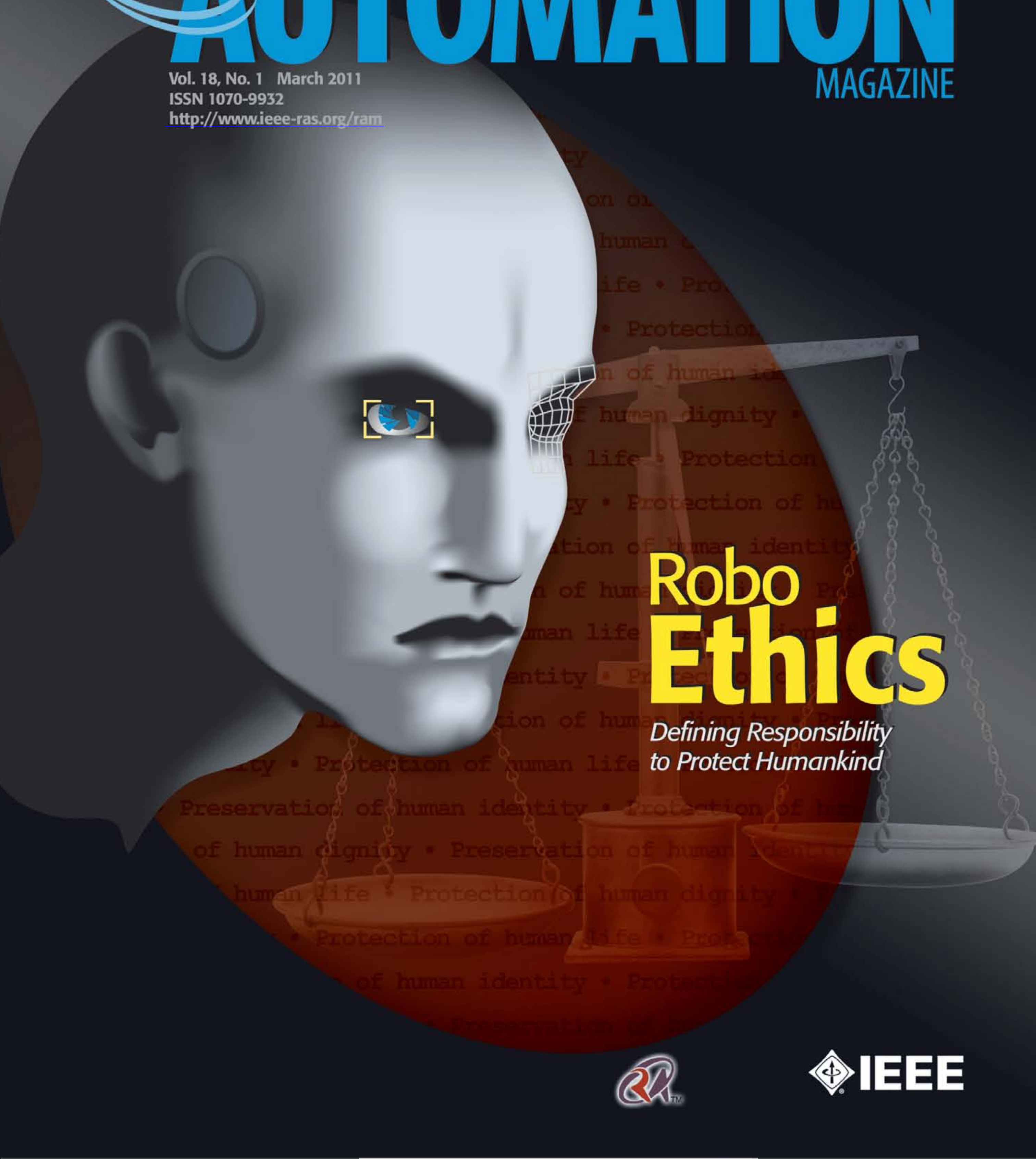
IEEE ROBOTICS & AUTOMATION MAGAZINE

Vol. 18, No. 1 March 2011
ISSN 1070-9932
http://www.ieee-ras.org/ram



Robo Ethics

Defining Responsibility to Protect Humankind

IEEE ROBOTICS & AUTOMATION MAGAZINE

Vol. 18, No. 1 March 2011
ISSN 1070-9932
http://www.ieee-ras.org/ram

FEATURES

© iSTOCK PHOTO.COM/VLADIMIR NIKITIN

ON THE COVER
This special issue deals with the emerging debate on roboethics, the human ethics applied to robotics. This issue also gives priority to broader articles that provide cultural and philosophical direction to those approaching the subject for the first time.

ROBOT IMAGE: © ISTOCKPHOTO.COM/AVIVANOV,
SCALES: © PHOTODISC

Digital Object Identifier 10.1109/MRA.2010.940140

© KARLSRUHE INSTITUTE OF TECHNOLOGY

COLUMNS & DEPARTMENTS

IEEE ROBOTICS & AUTOMATION MAGAZINE

A Publication of the IEEE ROBOTICS AND AUTOMATION SOCIETY

Volume 18, No. 1 March 2011 ISSN 1070-9932 http://www.ieee-ras.org/ram

IEEE prohibits discrimination, harassment, and bullying. For more information, visit http://www.ieee.org/web/aboutus/whatis/policies/p9-26.html.

IEEE Robotics & Automation Magazine (ISSN 1070-9932) (IRAMEB) is published quarterly by the Institute of Electrical and Electronics Engineers, Inc. Headquarters: 3 Park Avenue, 17th Floor, New York, NY 10016-5997 USA, Telephone: +1 212 419 7900. Responsibility for the content rests upon the authors and not upon the IEEE, the Society or its members. IEEE Service Center (for orders, subscriptions, address changes): 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855 USA. Telephone: +1 732 981 0060. Individual copies: IEEE members $20.00 (first copy only), nonmembers $96.00 per copy. Subscription rates: Annual subscription rates included in IEEE Robotics and Automation Society member dues. Subscription rates available on request.

Periodicals postage paid at New York and additional mailing offices. Postmaster: Send address changes to IEEE Robotics & Automation Magazine, IEEE, 445 Hoes Lane, Piscataway, NJ 08854 USA. Canadian GST #125634188

PRINTED IN THE U.S.A.

IEEE


KUKA
Desktop mobile manipulator for education and research
KUKA youBot
www.kuka-youbot.com
www.youbot-store.com
KEY FEATURES:
omnidirectional mobile platform
5-DOF manipulator
two-finger gripper
real-time EtherCAT communication
open interfaces
arm and platform can be used independently

FROM THE EDITOR'S DESK

New Look for *IEEE Robotics and Automation Magazine*

By Peter Corke

Welcome to the first issue of *IEEE Robotics & Automation Magazine* for the year 2011. I hope you would have already noticed some significant changes. We have a new cover style and internal layout. There were several motivations for doing this. First, we've had the current layout for many years now and it's a little dated. The move to digital delivery was an extra impetus for the change, and the layout has been designed to look great on paper and screen. Finally, the print production technology has improved, which allows us to do these great-looking covers (which is rather ironic as we start the move away from print).

Ethics is a topic that I believe is of great importance to us as a technical community.

Welcome also to the ethics issue. Ethics is a topic that I believe is of great importance to us as a technical community and will become more so as robots become increasingly pervasive in society. Don't ignore it. Please take the time to read these articles and debate them with your colleagues, professors, or students. An ethics issue was the top among my personal list of things I wanted to achieve with this magazine when I took over last year. If you feel strongly about any of these articles, agreement or disagreement, please share your views and arguments with our readers. E-mail me with a "Letters to the Editor" style piece, and we will publish it in an upcoming issue and/or on our Web site.

The campaign for a more *magaziney* magazine continues, and there are some changes in the mix of manuscripts that are coming in. I'd be particularly interested to see articles about real-world experiments. Many of us are involved in field trips or tests, and these often require hard work and are conducted in challenging circumstances and environments. The learnings from these exercises can also be extreme, not just about our technologies but rather on the social and human aspects of project teams and robot/human interaction. Such stories could be the basis of great articles.

The finances of our Society are complex and not widely understood. Quite a bit of the complexity is due to the way in which the IEEE operates as a not-for-profit organization. To get a straight answer, I've gone to the top and asked Bill Hamel and Xiaoping Yun, our vice president of financial activities and our treasurer, respectively, to write a short article, see page 8, explaining how it works.

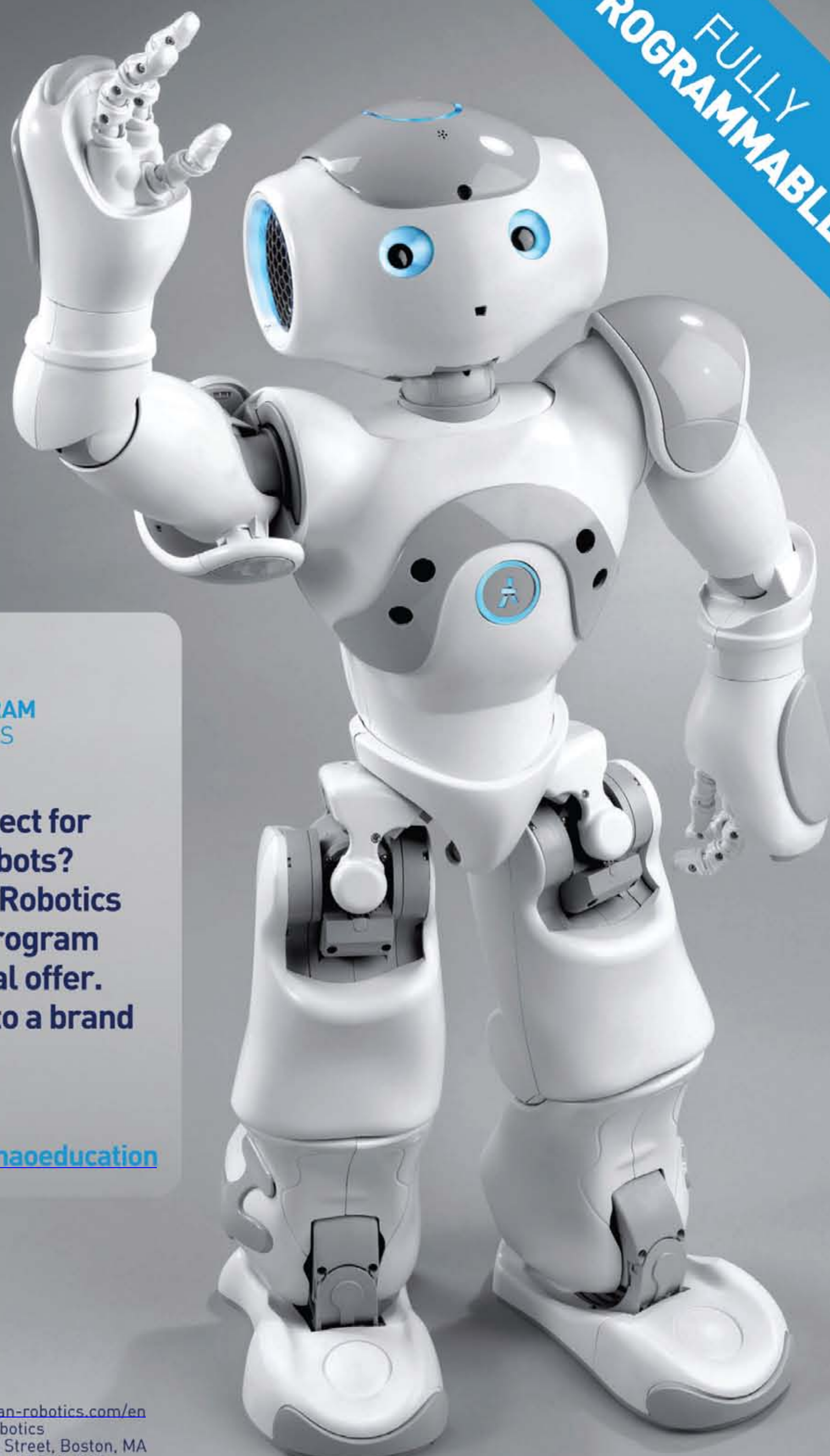
The tutorial in this issue is part one of two on the topic of "Motion Planning" written by Steve LaValle. If you have ideas for a tutorial that you would like to write, please contact me directly.

Right now the magazine is without a book review editor, and this is something I would like to rectify. There are many good titles coming out in robotics and automation as well as our sister fields of computer vision, control, mechanics, and signal processing. There is opportunity for new ways to do a books column, and it doesn't have to be just a matter of writing a major review or begging others to do reviews. Perhaps short reviews of many books or rolling reviews of current titles and their respective strengths. Maybe we can take it online and have our colleagues review the books that they use. If you have an interest in giving a new look to the books column, please contact me directly. The "Turning Point" column has a different interviewer for this issue. Eric Guizzo, *IEEE Spectrum*'s Automaton blogger, follows up his recent *IEEE Spectrum* article on telepresence robots by interviewing Ken Goldberg.

As always, I'm happy to hear ideas and opinions about this magazine. You can contact me by e-mail or come and find me at the IEEE International Conference on Robotics and Automation in Shanghai.

Enjoy the issue.

Digital Object Identifier 10.1109/MRA.2010.940141
Date of publication: 14 April 2011



Remarkable Increase in Student Membership

By Kazuhiro Kosuge, Tohoku University

One year has already passed since I started to serve as the president of the IEEE Robotics and Automation Society (RAS). I am writing this column in December, and I am very happy to report that our Society has been continuously growing. By the end of January 2011, our Society had a total of 8,102 members, which is an increase of 12.9% compared to that in January 2010, the highest percentage membership increase among the IEEE Societies. Student membership showed a remarkable increase of 29.3% compared to that in the previous year. Only two out of 38 Societies of the IEEE, RAS and Power & Energy, had an increase of more than 9%.

Digital Object Identifier 10.1109/MRA.2010.940143
Date of publication: 14 April 2011

We have more good news. The 2011 IEEE Fellows have been announced, and 15 of the candidates evaluated by the RAS were elected along with the three RAS members who were evaluated by other Societies. See the list of new Fellows with their citations in the "Society News" column of this issue (p. 100). Congratulations for their achievements and elevation to the IEEE Fellow.

In 2010, we decided to reduce the Society membership fee for 2011 and, at the same time, hire a new staff member at the IEEE headquarters in Piscataway. We hope that the reduced membership fee will encourage more researchers/practitioners to join our Society, especially with the new IEEE e-membership for the developing nations. The new staff member is going to work with Rosalyn Snyder, our staff administrator in North Carolina, to make it possible to handle the increased number of activities and services to the members of our Society. Actually, the RAS and IEEE have initiated more activities, and it has become impossible to take care of all the activities by a single staff person.

I regret to announce that we had some bad news in 2010. We experienced serious plagiarism cases in our past and future flagship conferences. The IEEE will take necessary actions on the basis of the IEEE ethics policies. We are not police officers. We do not want to allow this to happen again. To prevent the plagiarism, we are planning to implement the so-called Cross-Check system in the submission system of conference papers. We hope in future this will help reduce the number of cases of plagiarism.

In this issue, we have a report from our Vice President for Financial Activities William Hamel and our treasurer Xiaoping Yun, which clearly explains the details of RAS and IEEE finances, which many have wondered about. We also have (in our "Society News" column) a report from the Vice President for Member Activities Stefano Stramigioli on the recent Member Activities Board news. The "Industrial Activities" column features a report on the RAS standards activities by the Associate Vice President for Industrial Activities Raj Madhavan.

I wish all of you a happy, productive, and prosperous year in 2011.

Income and Expense Streams of RAS

By William R. Hamel, Vice President for Financial Activities
and Xiaoping Yun, Treasurer

The financial activities of the IEEE Robotics and Automation Society (RAS) are one of the most important, and yet often misunderstood, aspects of our organization. Financial details are at the very heart of the Society's identity and stature with respect to RAS members and the overall IEEE. Sound financial operations assure that we can offer our members world-class publications and conferences as well as a host of other benefits including awards and special events. Yet, when some hear about the large amounts of money associated with the Society reserves and the costs of many of the Society's products, they question things such as the IEEE's restrictions on "our" money held in reserves and the magnitude of their costs, such as specific conference registrations fees and even IEEE membership fees. This often begs the question: Are we taking too much money from our members and using it in ways that don't benefit them to the degree it should? The purpose of this article is to summarize the Society's financial mechanics and facts, with the hope of giving you a greater awareness of how things work. We should note that we have taken some license in our discussion to simplify things in the interest of focusing on the main issues.

From a financial perspective, the RAS operates like a business unit within the IEEE.

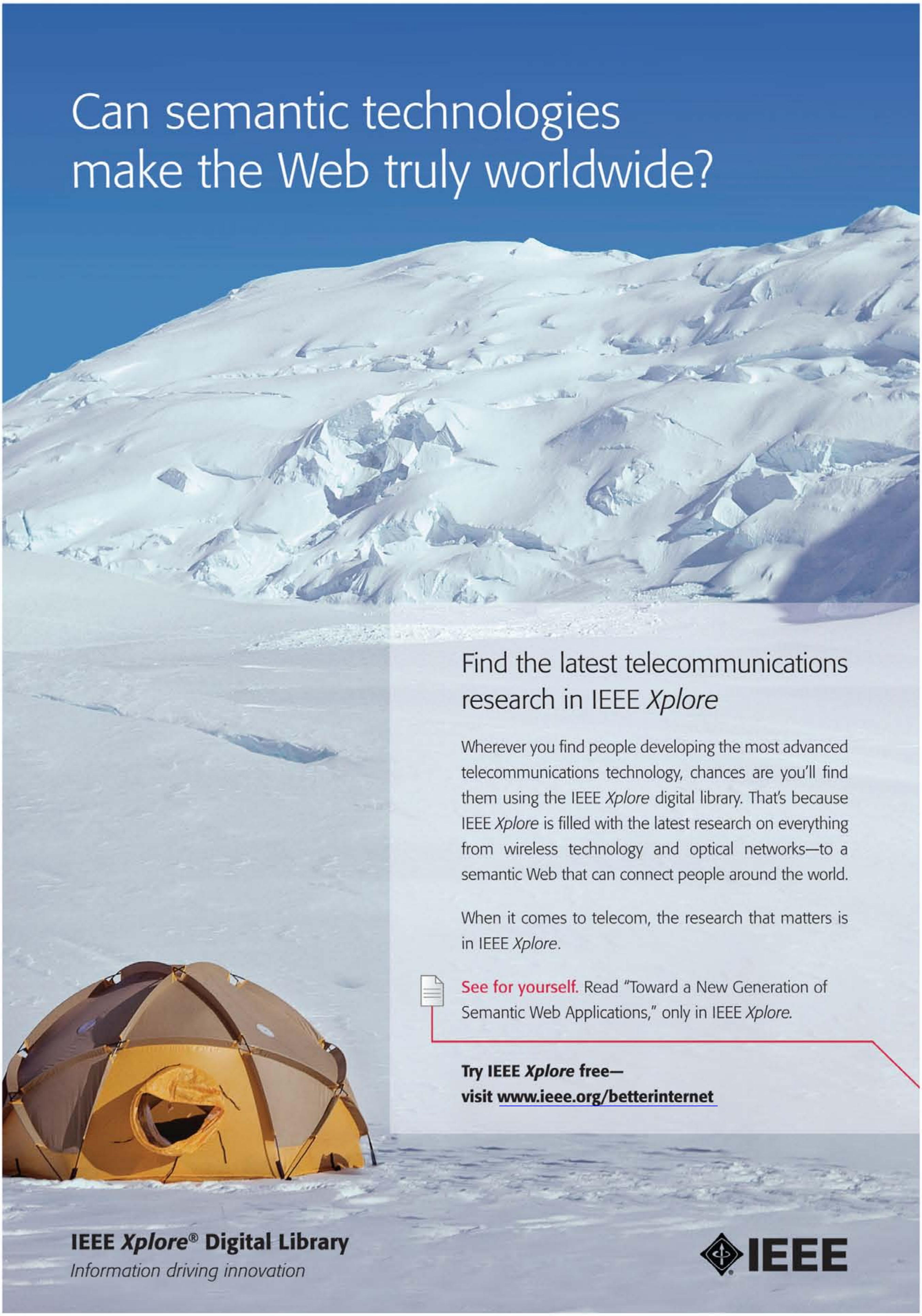
Digital Object Identifier 10.1109/MRA.2010.940144
Date of publication: 14 April 2011

The Basic Financial Model

From a financial perspective, the RAS operates like a business unit within the IEEE. We are expected to manage our income and expenses in ways that are beneficial to our members and the IEEE while being consistent with the IEEE policies and its not-for-profit tax status. The IEEE is the largest engineering professional society in the world, and, often, their policies, procedures, and rules seem endless and overpowering; yet, the RAS benefits greatly from the internationally recognized IEEE brand. We believe that the IEEE actually gives its subunits, 38 Societies and seven councils, more autonomy than other major technical societies. We are allowed to define and manage our business activities in ways we judge are in the best interests of our members.

The RAS budgeting and budget-control processes center on the balancing of our income and expense streams. Our income stream is a by-product of the Society activities and is made up of certain basic sources. These include membership fees, conference surpluses, earnings from our journals and conference proceedings, interest earnings from the investments made with our Society's financial reserves, and magazine advertisements. The expense streams are composed of the costs of various Society activities such as producing our publications, including editorial expenses, delivering conferences, providing financial support to member activities (e.g., RAS local Chapter grants, student/author travel support to attend the IEEE International Conference on Robotics and Automation, and the Distinguished Lecturer Program), presenting awards, operating the Society administration through the officers and administrative committee, our portion of the IEEE overhead and infrastructure costs, and any RAS-dedicated staff support costs.

In a given year, the expenses may be less or greater than the income. If there is a deficit, the Society reserves are used to balance the budget. If there is a surplus, the funds are moved to the RAS financial reserves. As any member would hope, the Societies are monitored very closely by the IEEE Technical Activities Board (TAB) with regard to deficit spending. A host of additional requirements and constraints become operative if a particular Society becomes significant in the red. The IEEE financial reserves are to a great extent the summation of the Society reserves. This means that the financial integrity of the IEEE is connected directly to the financial integrity of its Societies. As a result, the IEEE manages these reserves carefully. In fact, when the recent stock market collapse reduced the overall IEEE reserve to a dangerously low magnitude, a moratorium was placed on the use of reserve funds. This moratorium was rescinded for the 2011 budget year. Not only do yearly surpluses but also any interest income from investments made with the actual reserve funds go into the Society reserves. When the stock market is strong, the RAS interest income from reserves investments can be significant. There are special rules that TAB has established regarding how societies may access and use the financial reserve funds. Basically, reserve funds can only be used for defined initiatives that are submitted and approved during the budget-development process each year. In addition, a maximum of


Can semantic technologies
make the Web truly worldwide?
Find the latest telecommunications research in IEEE *Xplore*
Wherever you find people developing the most advanced telecommunications technology, chances are you'll find them using the IEEE *Xplore* digital library. That's because IEEE *Xplore* is filled with the latest research on everything from wireless technology and optical networks—to a semantic Web that can connect people around the world.
When it comes to telecom, the research that matters is in IEEE *Xplore*.
See for yourself. Read "Toward a New Generation of Semantic Web Applications," only in IEEE *Xplore*.
Try IEEE *Xplore* free—
visit www.ieee.org/betterinternet
IEEE *Xplore*® Digital Library
Information driving innovation
IEEE

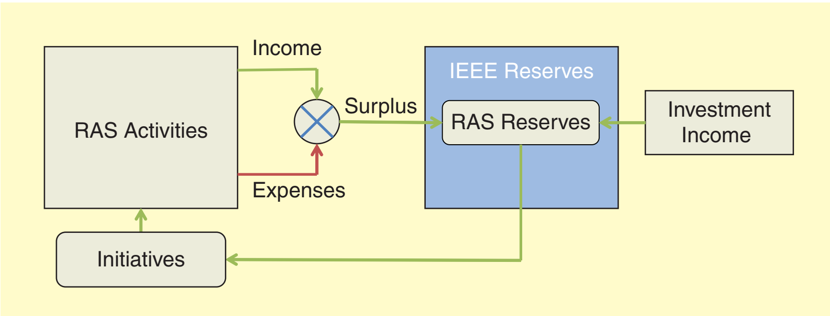


Figure 1. RAS financial mechanics.

3% of the reserve fund may be extracted for specific short-term initiatives in a given year. From an engineering perspective, the Society's financial reserve looks like an accumulator that has an input line with a check valve that restricts the outflow. Money can flow into the reserves at a much larger rate than it can flow out. One of the biggest challenges that the Society leadership faces is controlling the budget such that the annual surpluses are as small as possible, which, in turn, means that the funds trapped in the reserves are as small as possible. They must also be as skillful as possible in using the initiatives process to pull funds from the reserves for projects/activities that benefit the Society membership. Figure 1 provides a schematic representation of the basic financial architecture.

When the stock market is strong, the RAS interest income from reserves investments can be significant.

A final point regarding IEEE oversight is that from time to time some of the Societies/Councils may find themselves in less than favorable financial positions. To protect the overall IEEE and the financially sound societies/councils, the TAB Financial Committee maintains a watch list, which is part of an active process to closely monitor their activities to assure that potential financial problems are contained. A society/council may end up on this list if they run, or plan, a deficit budget in two years out of a three-year window or if the Society/council does not maintain minimum levels of reserves for two years in a three-year window. There is a clear degree of negative recognition for a Society to be on this list. A major goal of your leadership is to not be placed on this list, which could happen should we experience budget deficits for more than two years out of three. As we seek to keep our surpluses low, such that we maximize the use of funds for our membership, we also increase the possibility of running a deficit. This is a major challenge because we do not have direct control over large fractions of our income streams, particularly, the earnings from publications and conferences. Within the RAS Financial Activities Board, we work hard with the other boards to understand all the details such that we can successfully develop and execute budgets that result in low but positive surpluses.

Budget Year 2011 and Summary

Each year, we are required to submit a detailed budget to the IEEE that is approved by TAB. As discussed above, we are expected to operate within the approved budget, within a reasonable margin of error, as the year unfolds. The RAS is in effect a sizable business, and the 2011 budget is based on a total income of US$3.788 million, with the expenses projected to be US$3.595 million. This means that our projected positive surplus is US$194,000. By the way, this includes significant reductions in membership dues and member publications fees, which in effect reduces income, that were initiated by President Kosuge and reported earlier. The 2011 budget surplus is a bit larger than our initial target mainly due to improved economic forecast that results in larger projected income from publications and conference proceedings. In the last several years, the budget surpluses were −US$789,000, −US$462,000, −US$425,000, −US$591,000, and US$781,000 in 2006–2010, respectively. A negative budget surplus means that efforts were made to extract up to 3% of reserve funds to support the Society activities.

Probably the largest uncertainty in the budgeting process is associated with the income generated from our conference and journal publications. Such earnings are dominated by returns from institutional/business downloads associated with the IEEE *Xplore* digital library. It is very difficult to predict, other than prior history, what the magnitudes of these earning will be in future years. In 2011, more than a total of US$1 million is projected, and you may find it surprising to know that conference publications' earnings for proceedings that are made available on IEEE *Xplore* are expected to be US$777,000!

In summary, the RAS budget, like any budget, is a balancing act that requires detailed understanding of the income and expense streams combined with active controls and diligence throughout the year to assure that deficit spending is avoided. Most importantly, this must be done with the overarching goal of striving to use the Society funds/resources to the maximum benefit of the members. Within the Financial Activities Board and in conjunction with the other officers and boards, we are dedicated to doing just that.

IEEE Robotics & Automation Magazine | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

Imitating Ourselves in Silicon

By Jeanne Dietsch

Humans have strived to create life from inanimate matter since time immemorial: from cave drawings to sculpted marble to electronic android. Recently, biomimetic robotics has made enormous advances, but the Holy Grail of reverse engineering, the human brain, skitters further away the closer we approach.

Reverse Engineering the Human Cosmos

Despite claims that the scientists will reverse engineer the human brain by 2030 [1], the experimental fluorescence of mouse brain synapses by Smith and Micheva [2] suggests that this feat may be 1,000 times more complex than was previously thought. Smith and Micheva developed a technique called *array tomography*, a high-resolution photography of jellyfish molecules that bind to different proteins and glow in different colors. Array tomography, for the first time, enables the scientists to count and categorize a snapshot of the 125 trillion or so synapses present in the cerebral cortex. It also revealed the jaw-dropping discovery that, far from being a binary device as previously suspected, each synapse resembles a microprocessor with as many as 1,000 switches. Hoping to map the human brain and recreate neurologically accurate androids, the researchers might need to extend their timelines a few decades as we investigate these switches further.

Figure 1. ARMAR-III learns about its environment through actions just as infants do. (Photo courtesy of Karlsruhe Institute of Technology.)

Mapping Cognitive and Perceptual Processes

Less complex biomimetic and biologically based systems are making significant strides forward. For instance, the European Commission Cognition Unit, led by Tamim Asfour of the Karlsruhe Institute of Technology, has developed ARMAR-III, which actively investigates its environment. In fact, entities therein only become semantically useful objects through the actions the robot performs on them [3] (Figures 1 and 2). For instance, ARMAR-III learns a cereal box's characteristics, such as lightweight and source of cereal, by lifting and turning the open box. It can compare these associated object-action complexes (OACs) by performing the same actions on a book or a block to differentiate the classes of objects. Mentors can teach ARMAR to relate specific OACs with a word, such as heavy or light, similar to the way the infants learn language. The Paco-Plus team plans to teach ARMAR to generalize, extrapolate, learn grammar, and refine the teaching methods based on goals.

Visitors to Heartland Robotics recently revealed a few more details of Rodney Brooks' dual-armed, compliant coworker robots [4]. The humanoid torsos with camera heads will be rolled into place, plugged in, and taught motion sequences without programming. Brooks states that the camera heads will offer

Digital Object Identifier 10.1109/MRA.2010.940145
Date of publication: 14 April 2011

the software developers the opportunity to use machine vision for quality control as the robot works, but it is difficult to believe that they will not deploy some of the perception and learning algorithms Brooks has developed over the decades.

Other perception researchers at the University of California, Santa Barbara, have found that the people's actions generate not only visual but also proprioceptive maps [5]. Bernier and Grafton's study is one of the first to suggest that even sighted people use proprioceptive mapping when left in darkness or with nonvisual objects. Via magnetic resonance imaging and repetition suppression, they found that people can switch instantly back and forth across independent visual and proprioceptive maps.

Another investigator of biological vision, Baluch of the University of Southern California, has put live locusts onto a mobile platform with a control algorithm that fuses information from multiple sensors to guide the robot [6]. "The locust's visual system responds vigorously to looming stimuli and the threat of imminent collision. Such stimuli evoke robust responses in motor neurons enabling evasive behaviors," states Baluch.

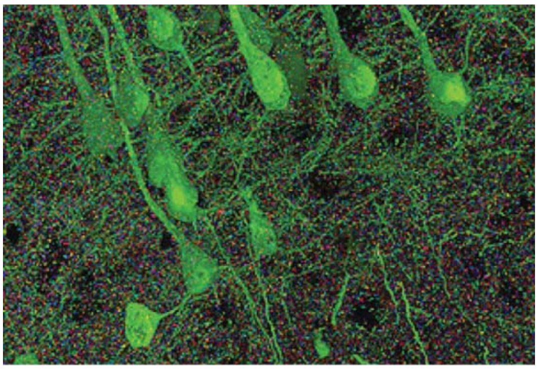


Figure 2. Scientists attempting to recreate the human brain face, stunning new discoveries of its complexity. (Photo courtesy of Stanford University.)

Figure 3. Pelican navigates autonomously over and under obstacles. (Photo courtesy of SEAS, University of Pennsylvania.)

Evolving Biomimetic Forms and Functions

Biomimetics has entered the industrial realm with pneumatic and hydraulic robots, manipulators, and graspers from companies such as Festo Corporation [7] and Inmótx [8]. Festo highlights several biomimetic prototypes including the elephant trunk-inspired Bionic Handling Assistant, the flying, jellyfish-like AirJelly, and the increasingly humanlike AirArm. Inmótx develops application-specific end effectors based on octopuslike suction and release that are used to handle previously impossible items from apples to chicken to packages of pasta.

Flying robots, like AirJelly, will be able to navigate more intelligently as a result of the research done by Michael of the University of Pennsylvania with Pelican in conjunction with Army Research Laboratory [9]. Pelican weighs less than 1.5 kg, yet has the computing power and laser camera sensing to navigate through a three-dimensional map and over and under obstacles (Figure 3).

Watching the progression of silicon-based systems in imitation of and even integration with carbon-based life forms is fascinating without a doubt. The ethical implications of such work will challenge the world's legal and cultural foundations as we seek to redefine which beings are created equal and which are not.

References

[1] R. Kurzweil, *The Singularity Is Near*. New York: Viking, 2005.

[2] K. D. Micheva, B. Busse, N. C. Weiler, N. O. Rourke, and S. J. Smith, "Single-synapse analysis of a diverse synapse population: Proteomic imaging methods and markers," *Neuron*, vol. 68, no. 4, pp. 639–653, Nov. 18, 2010.

[3] (2011, Sept. 3). PacoPlus Web Site [Online]. Available: http://pacoplus.org/

[4] (2011, Dec. 6). Heartland wants its robot to give manufacturing a lift. *Boston Globe* [Online]. Available: http://www.boston.com/business/technology/articles/2010/12/06/heartland_wants_its_robot_to_give_manufacturing_a_lift/

[5] P.-M. Bernier and S. T. Grafton, "Human posterior parietal cortex flexibly determines reference frames for reaching based on sensory context," *Neuron*, vol. 68, no. 4, p. 776, 2010.

[6] F. Baluch, "Putting locust brains to work: Robotic collision avoidance using locust neurons," in *Proc. 5th Annu. Neuroscience Graduate Student Symp.*, USC Davidson Conference Center, Jan. 22, 2010.

[7] (2011, Sept. 3). Festo Web Site [Online]. Available: http://www.festo.com/cms/en_corp/9617.htm

[8] (2011, July 12). Inmotx Web Site [Online]. Available: http://www.inmotx.com

[9] U.S. Army News, "Robot 'thinks' its way through obstacle course," in *Proc. Army Science Conf.* [Online]. Available: http://www.army.mil/-news/2010/12/01/48864-robot-thinks-its-way-through-obstacle-course-at-army-science-conference/

Jeanne Dietsch is the vice president of Emerging Technologies at Adept Technology Inc. As a cofounder of MobileRobots, purchased in 2010 by Adept, she led the development of motivity autonomous robot-operating system and other foundational technologies used in autonomous mobile robots and automated guided vehicles. She blogs for *IEEE Spectrum* and serves on the Industrial Activities Board (IAB) of the IEEE Robotics & Automation Society as well as the editorial boards of various robotics journals. She helped in founding the Robotics Technology Consortium.

COMPETITIONS

An Overview of RoboGames

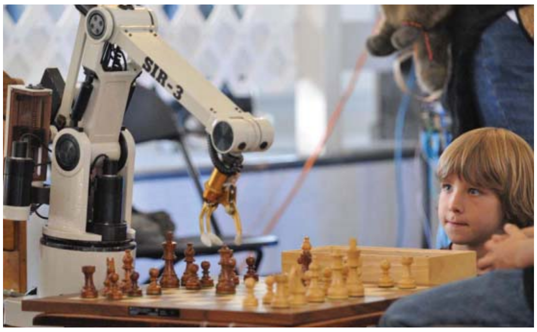
By David Calkins

This column features an overview of RoboGames, officially recognized as the world's largest robotics competition by *The Guinness Book of Records*. It is an extremely wide-ranging set of competitive robot events, open to everyone. It is annually held in the United States and attracts competitors from all over the world.

In 2009, Rodi Hartono, a 21-year-old undergraduate from Indonesia, went to America for a week, with a singular goal in mind: winning a gold medal. He returned home as a star and was even personally congratulated by the Indonesian President Susilo Yudhoyono, who presented Rodi with a lifetime scholarship through postdoctorate studies. However, Rodi was not an athlete; he was a robot builder who competed at the international RoboGames. Every year, hundreds of robot builders from all over the world compete with other students, professionals, and dedicated amateurs, all with a dream to win gold medals. The teams range from hobbyists to IEEE organizations (such as the University of Waterloo's IEEE Humanoid Team).

RoboGames was founded in 2004 as an exercise in cross-pollination. Too often, robot competitions specialize in a single discipline (electrical and mechanical engineering, computer science, sensor technology, etc.), and participants do not get exposed to the other areas of robotics. Rarely does a sumo robot builder work on robot soccer or would vision experts try their hand at mechanical engineering. RoboGames provides an opportunity for the participants to see what exciting things are going on outside their specialty in an event created to celebrate all robotic disciplines.

RoboGames has achieved wide acclaim (recognized by Guinness as the world's largest robot competition) but was founded with four simple goals:



A robot playing chess at the 2009 RoboGames. (Photo courtesy of Alan Musselman.)

1) *Get robot builders, amateur and professional alike, to grow beyond their specialization.* Robot soccer, for example, primarily tends to be a programming exercise. The often-maligned combat builders are brilliant mechanical and electrical engineers but rarely use sensors. Misunderstandings can be resolved and more breakthroughs achieved when various robot subspecialists see each other's work first hand.
2) *Let everyone compete.* Anyone, regardless of his/her age, affiliation, country of origin, gender, or academic discipline (or lack thereof), can compete. Many events limit who can compete: they might be offered only to high school or university teams or have entry fees in excess of US$5,000. RoboGames expands on these forums and welcomes everyone with registration fees lower than US$250 (several events are free).
3) *Expose the general public to robots they would not otherwise see.* Although events such as RoboCup and FIRA lead to great developments, the general public rarely attends them. Robots should be seen by everyone, not just other researchers. At RoboGames, the public is drawn in by the combat robots but is then successfully exposed to the other competitions.
4) *Give robot builders the recognition they deserve.* RoboGames highlights the achievements of its participants through widespread mainstream media exposure, continued promotion to the general public, and year-round advocacy. Several medalists have met their nations' leaders owing to their success at RoboGames.

To meet these goals and serve as many types of robot disciplines and builders as possible, RoboGames offers more than 70 different robot competitions:

- *Humanoids:* This is the fastest growing segment with competitions including autonomous freestyle, biped race, weightlifting, soccer, stair climbing, obstacle course, kung-fu, and several others.
- *Combat:* This is an audience-favorite game, pitting mechanical and electrical engineers (many design their own speed controllers and drive systems) against each other, while still fostering great sportsmanship. Combat has proven to be a highly effective gateway to STEM education and other events.
- *Soccer:* While RoboGames cannot offer every robot soccer variant, MiroSot and 3:3 humanoids attract both contestants and a large audience.
- *Sumo:* The oldest robot sport now has six weight classes starting from just 25 g up to 3 kg. Most matches end within 10 s, as the autonomous bots quickly force their opponent out of the ring.

Digital Object Identifier 10.1109/MRA.2010.940146
Date of publication: 14 April 2011

IEEE Robotics & Automation Magazine
Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

- *Hockey:* One of the tough things about robots is their agility at high speed. Hockey bots not only get the competitors building more agile frames but the crowds go crazy too.
- *Tetsujin:* Man meets machine when contestants strap on a body suit like Sigourney Weaver in *Aliens* to lift weights and walk without their leg muscles. Monty Reed of Washington builds suits that allow him to stand up and walk without using his legs. Who needs wheelchairs?
- *Autonomous navigation:* Full-sized DARPA Grand Challenge autonomous cars are prohibitively expensive for most competitors, so RoboMagellan scales it down to R/C car size but with autonomous vision and navigation systems to find GPS waypoints while avoiding obstacles. The NatCar event gets contestants following complex lines and paths.
- *ArtBots:* The form follows the function, and the art bots category tries to encourage robot builders to make their bots much more aesthetically available. Static, kinetic, drawing, musical, and bartending robots add a new dimension.
- *BEAM:* Analog robots are not forgotten among all the digitally programmed bots. The three-BEAM challenges let analog engineers put their skills to the test.
- *Open:* The "everything else" categories such as micromouse, line followers, firefighting, walker challenge, two-wheeled balancers, and hexapods fill out the events. There is also a "best of show" category for robots that do not fit into any other class.
- *Jr. League:* To encourage kids to start building and get involved in STEM education, RoboGames offers ten no-cost kids events.

The goal of RoboGames of cross-pollination is already working. Humanoid builders come to know about better sensors from sumo teams. Sumo builders learn to strengthen their robots from combat engineers. Combat builders have gone from building robots in their spare time to founding companies that produce hardened, teleoperated bots for police and military buyers. Children who started out building LEGO robots have grown into engineers working at robotics companies (such as 19-year-old Tony Pratkanis, five-time gold-medal winner and now a Willow Garage employee).

The goal of RoboGames is to welcome more robot enthusiasts into the event as competitors, mentors, builders, and supporters, with a steady influx of enthusiastic, inspiring, and talented people. Taking part in competitions is great, but the hook is the lasting and valuable relationships forged between the attendees who met at RoboGames. The 2011 event is from 15 to 17 April as part of National Robotics Week. With so many events, there is something for every robot. So start planning now. More information is available at http://robogames.net.

INDUSTRIAL ACTIVITIES

RAS Standing Committee for Standards Activities—An Update on Recent Activities

By Raj Madhavan

Standards are crucial for driving industry innovation and technology transfer. Benchmarking and standardization are vital to the development of robotic and automation systems in already-established application areas and are critical to wider (including societal) acceptance of emerging technologies. It is widely accepted within the robotics and automation community that leaving emerging robotic technologies to proliferate in an unguided direction comes with a high price: synergistic opportunities remain unrealized and a lack of cohesion in the community hinders the progress in many domains such as manufacturing, service, health care, and security, to name a few [1], [2]. For an in-depth discussion of the existing standards and worldwide efforts, an interested reader is referred to [3].

Digital Object Identifier 10.1109/MRA.2010.940148
Date of publication: 14 April 2011

The Standing Committee for Standards Activities (SCSA) under the Industrial Activities Board (IAB) of the IEEE Robotics and Automation Society (RAS) is working together with the research and industrial communities and other standards developing organizations to help develop standards for robotics and automation. The scope of the activities of the RAS-SCSA is to formally adopt and confirm best practices in robotics and automation as standards. Within this scope, the RAS-SCSA is pursuing the following objectives [4]:

- promote common measures and definitions in robotics and automation
- promote measurability and comparability of robotics and automation technology
- promote integrability, portability, and reusability of robotics and automation technology.

Some of the previous work carried out by the Standards Committee can be found in [5] and [6]. At the 2010 International Conference on Robotics and Automation (ICRA 2010) held in Anchorage, Alaska, the SCSA hosted two meetings. One of these meetings was organized by the IEEE Standards Association (IEEE-SA) to better understand the procedures involved in the standards process. The second meeting was part of the SCSA's regular meeting series, which delved into identifying the suitable areas for standardization that are both crucial and achievable in the short term. These meetings served as an excellent forum to discuss and exchange ideas from professionals working on various aspects of robotics and automation and many with previous experiences in standards development. Two follow-up meetings were held at the 2010 International Conference on Intelligent Robots and Systems (IROS 2010) in Taipei, Taiwan, to further discuss the scope of the study groups (SGs) and to develop a timeline for the standards-defining activities.

On the basis of the discussions stemming from these meetings and with input and consensus from participating members, two SGs have been formed in the following topic areas.

Map Data Representation and Fundamental Data Types

The objectives of this SG are to study the existing map data representation(s) and discuss how to represent, encode, and exchange map data for robot navigation via standardized data-exchange formats to enable efficient and proper use of robot software frameworks.

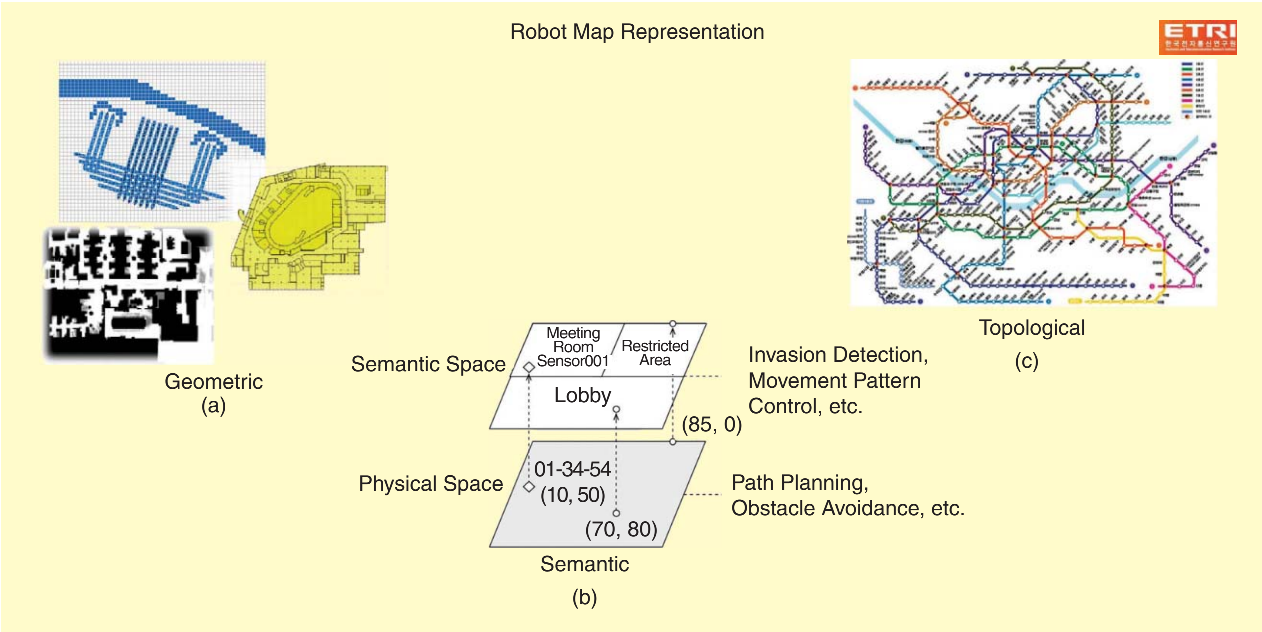
Various applications are emerging in the robotics industry, including security, transportation, and elderly care services, to mention a few. Mapping, a crucial process in robot navigation, is a procedure to interpret sensor readings about the surrounding environment into machine-understandable environmental features. It has a long history of research in the robotics field and has strong industry applications such as autonomous vacuum-cleaning robots and on-road navigation. Regardless of the mapping algorithm employed, a resulting map data should be interoperable for other robots or robotic service providers to perform or control the navigation function. Accordingly, this SG aims at developing a consensus on the needs for common representation for robot map data, including geometric, topological, or semantic maps. The SG will investigate the potential for standardizing fundamental data types for mapping and examine the existing map data representation coming from other technical organizations such as Open Geospatial Consortium. By doing so, the SG will set up a long-term road map for robotic mapping and discuss how to represent, encode, and exchange map data for robot navigation.

In addition, as part of the task of producing a standard for mapping in robotics, there is the task of identifying the data types involved, including those reaching down to the fundamental level. This stems from the long difficulty in robotics of matching data types used between the software existing in separate projects and the need for mapping systems to exchange data with the other robotics software. To date, many robot software frameworks have supplied, in some way, a collection of data types for use with the framework. Player, for example, specified the player abstract device interface. Frameworks as recent as OpenRTM-aist and ROS have followed this course eventually in an effort to prevent fragmentation. To promote interoperability between frameworks exchanging map data, it is necessary for the data types involved to be defined down to a fundamental level. Defining the data types involved down to such a low level will not solve the interoperability problem and ensure that map implementations can exchange data with each other. There are many other issues involved, such as transport protocols and calling conventions. It will, however, ensure that mapping implementations that overcome these other issues will be able to understand the core data involved. The SG will, as part of its investigation of standard map data types, also investigate the potential for standardizing the lower-level data types. Its key goals will be to investigate how low the data types should go and whether there is any real benefit of standardizing the lower data types at all, versus other options such as data-interchange protocols. The information produced will update about the standardization process.

Glossary/Ontology for Robotics and Automation

The objective of this SG is to identify, develop, and document the salient terms and their definitions so that they can serve as a common reference for the robotics and automation community.

There are many advantages of creating such a glossary/ontology, including 1) ensuring a common understanding among the members of the community, which helps to ensure timely decision making and minimizes potential confusion and 2) facilitating more efficient data integration and transfer of information among systems. Unlike previous efforts that have attempted to perform similar activities in more specific domains, this SG will take a more all-encompassing approach, not only focusing on terms that relate to the more traditional mobile robotics domain



(a) Some examples of metric map, (b) conceptual illustration of semantic mapping between physical space and the corresponding semantic description, and (c) (pure) topological map (subway lines of Seoul).

(e.g., service robotics, health-care robotics, and military robotics) but also extending it into the automation field that could include terms related to domains such as automated manufacturing shop floor. The exact scope of these efforts and the domains that will be addressed will be one of the first orders of business of this SG. Initial efforts will attempt to leverage previous and existing efforts that have tried to address parts of this issue, including the robots standards and reference architectures, autonomy levels for unmanned systems framework, and RobotWorx glossary. A core issue that will be addressed as part of this effort is how to represent the definition of the terms. Many approaches exist for representing knowledge at different levels of formality, including dictionaries and glossaries, database, and logic-based approaches. All offer their benefits and costs that need to be considered when determining the most appropriate representational approach. A related issue is how to most logically organize the terminology, especially in an area so large with many interrelationships.

It is anticipated that the symbiosis and interaction between these SGs will facilitate better discussions and rapid progress than what would be possible if these groups tried to address these problems in a stand-alone and isolated manner. The efforts of each of the SGs will culminate in the drafting of a proposal that will be widely circulated among the community for feedback and comments. The finalized SG document will, in turn, result in a project authorization report toward the formation of a working group (WG) that will work closely with the IEEE-SA. The WG document is expected to be available by April 2011. A full-day standards meeting will take place on the first workshop day of both ICRA 2011 and IROS 2011 to encourage participation of conference attendees in discussions and as a forum to bring together interested parties.

For 2010–2011, the author is serving as the chair of the Standards Committee under the IAB of the Society. Your suggestions and participation are most welcome. You can reach the author at raj.madhavan@ieee.org.

Acknowledgments

The author thanks Wonpil Yu (Electronics and Telecommunications Research Institute, Korea), Geoffrey Biggs (National Institute of Advanced Industrial Science and Technology, Japan), and Craig Schlenoff (National Institute of Standards and Technology, USA) for serving as the leads for the SGs and for their contribution to the article. Thanks are also due to the members who have actively participated in the standards meetings and have offered their comments. The author acknowledges the support of IAB Vice President Alex Zelinsky, RAS President Kazuhiro Kosuge, and the RAS AdCom.

References

[1] R. Madhavan, A. P. del Pobil, and E. Messina, "Performance evaluation and benchmarking of robotic and automation systems," *IEEE Robot. Automat. Mag.*, vol. 17, no. 1, Mar. 2010.

[2] E. Prassler, R. Bischoff, and H. Bruyninckx, "Its a long way to useful standards," *IEEE Robot. Automat. Mag.*, vol. 17, no. 3, pp. 18–19, Sept. 2010.

[3] R. Madhavan, R. Lakaemper, and T. Kalmar-Nagy, "Benchmarking and standardization of intelligent robotic systems," in *Proc. Int. Conf. Advanced Robotics (ICAR)*, Munich, Germany, June 2009, pp. 1–7.

[4] "Operating procedures and policies of the IEEE Robotics and Automation Standing Committee for standard activities," Apr 2007.

[5] S. Lee, E. Prassler, R. Tuokko, and H. Moradi, "The standardization and roadmapping initiatives by the industrial activity board," *IEEE Robot. Automat. Mag.*, vol. 12, no. 4, p. 98, Dec. 2005.

[6] E. Prassler and K. Nilsson, "1,001 robot architectures for 1,001 robots," *IEEE Robot. Automat. Mag.*, vol. 16, no. 1, p. 113, Mar. 2009.

IEEE Robotics & Automation Magazine
Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

Exponential Growth of ROS

By Steve Cousins

The open-source robot operating system (ROS) has been growing exponentially and has reached the critical mass. Every time we start to write this column, this fact worries us because, we fear, whatever we write now (three months before this magazine gets published) will be very old news by the time you read it. Nevertheless, in this edition, we'll report on the state of things in the ROS world, with the hope that at least we'll spark your interest in visiting http://ros.org to learn about the latest developments.

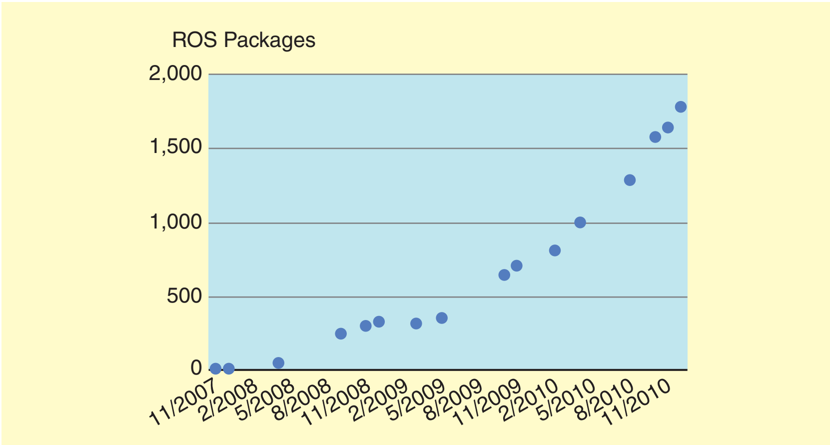
The growth of a worldwide community is little tricky to measure, but we have a lot of indicators. One measure of the community is the number of people contributing to it. Figure 1 shows the growth in the number of public repositories hosting freely available ROS code from the first contribution to ROS by the end of 2010.

Another important measure is how much code people are contributing, which we measure in unique ROS packages (Figure 2). Since November, the driver for ROS growth has been a new two-/three-dimensional (2-D/3-D) sensor from a company called *PrimeSense*, which has been made available to the mass market as Microsoft Kinect. This US$150 sensor delivers both 2-D camera data and 3-D point clouds at the same time, at a distance of about 0.5–3.5 m. These data are useful for applications ranging from autonomous navigation to 3-D model building and to novel new gesture-based interfaces for computers and robots. Kinect sold 2.5 million units in the first 25 days, which has immediately created a huge new potential user base for 3-D libraries.

The Kinect sensor has come at a great time for ROS. In addition to providing a much lower cost sensor for robotics, there is now an open-source library called *point cloud library* (PCL; http://pointclouds.org). PCL is designed to take advantage of the 3-D data that Kinect produces by providing state-of-the-art 3-D computer vision techniques such as surface reconstruction, segmentation, registration, feature detection, and so on. The PCL should form a perfect complement to OpenCV, which interacts mainly with the traditional cameras.

PCL has already been used in several innovative applications of the Kinect sensor. A researcher at the Massachusetts Institute of Technology used PCL to build a hand-detection algorithm, which he then used to create a virtual piano as well

Digital Object Identifier 10.1109/MRA.2010.940147
Date of publication: 14 April 2011



The number of ROS packages (basic code units) available in public repositories over time.

as a minority report interface. Research at Berkeley mounted Kinect to a quadrotor and used PCL's algorithms to detect the floor and obstacles. In short, having a low-cost sensor available has already enabled the community to take advantage of state-of-the-art software to perceive the world.

"All applicable software developed under the M3 program must be compatible with ROS: the ROS framework."

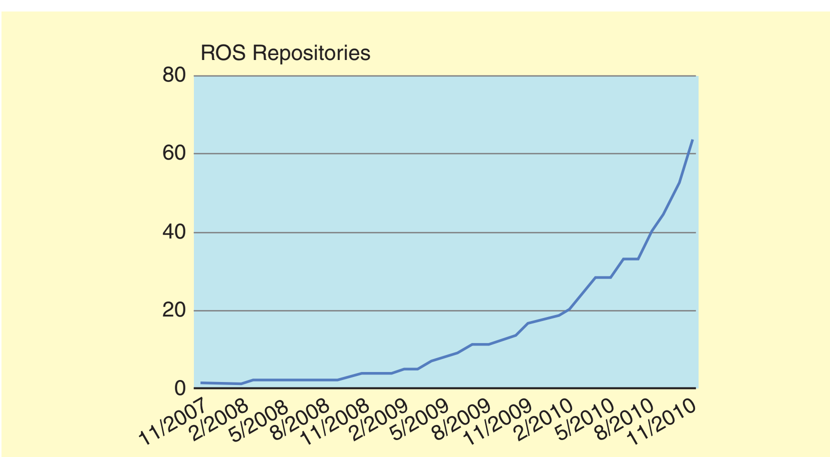
Another factor driving ROS adoption has been its integration with other open-source frameworks. For example, the researchers at K.U. Leuven have improved the integration of ROS with Orocos and now support the use of ROS packages within Orocos. Recently, the researchers at ETH Zurich integrated the ASEBA architecture for distributed control of mobile robots with ROS. An open-source award winner Geoffrey Biggs has provided a link between the ROS and OpenRTM middleware from AIST in Japan.

Outside of academia, commercial robotics packages are also joining the ROS bandwagon. SRI released much of its Karto mapping and navigation system as an open source, and it has provided a nice complement to other ROS tools. Gostai has made Urbi compatible with ROS and released the Urbi kernel as an open source. Urbiscript lets the users script the behaviors of the robot using primitives such as parallelism and event triggers. Finally, PrimeSense asked Willow Garage to join their OpenNI Foundation as a founding member to focus on robotics applications with ROS.

On the government side, ROS is now being advocated by two U.S. Defense Advanced Research Projects Agency (DARPA) mobile manipulation programs. The use of ROS is encouraged by the DARPA ARM program, and the BAA for the M3 program says, "all applicable software developed under the M3 program must be compatible with ROS: the ROS framework."

In addition to the robots with the Kinect sensors, ROS runs on well more than 50 robots, ranging from highly capable mobile manipulation platforms such as the PR2 or the Care-O-bot to inexpensive platforms such as the Lego NXT or the Neato XV-11. It runs on humanoids, autonomous cars, unmanned aerial vehicles, and autonomous underwater vehicles. You can see the whole list at http://www.ros.org/wiki/Robots. It is now possible to build a sub-US$1,000 robot using the Kinect sensor and other off-the-shelf components that will leverage ROS's strengths in navigation, 3-D processing, and community.

The next ROS release, named *Diamondback*, is coming soon (meaning that it will probably be out before you read this). In addition to a myriad of system improvements and the usual bug fixes, Diamondback will add a number of high-level capabilities to ROS. The SMACH executive that we discussed in the last column of "ROS Topics" will be standard in Diamondback. PCL will be integrated along an object-grasping pipeline and arm-navigation stack that can autonomously plan collision-free trajectories.



The number of ROS repositories corresponds roughly to the number of institutions contributing open-source packages to ROS.

Acknowledgment

I thank Ken Conley of Willow Garage for his help with this column.

Robotics & Automation Magazine | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

FROM THE GUEST EDITORS

Roboethics: Ethics Applied to Robotics

By Gianmarco Veruggio, Jorge Solis, and Machiel Van der Loos

This special issue deals with the emerging debate on roboethics, the human ethics applied to robotics. Is a specific ethic applied to robotics truly necessary? Or, conversely, are not the general principles of ethics adequate to answer many of the issues raised by our field's applications? In our opinion, and according to many roboticists and human scientists, many novel issues that emerge and many more that will show up in the immediate future, arising from the upcoming marketed robotics products, demand the development of new cultural and legal tools that can provide the crucial answers to the most sensitive questions.

The unfolding and emerging scenarios made possible by robotics are fascinating and unsettling at the same time. Suffice it to think that all machines, of any form and dimension and for any type of use, will be computerized, equipped with artificial intelligence and networked, to understand that everything we have seen to date—computers, video games, cellular phones, and Internet—is really only the dawn of the technological world that awaits us. For instance, in aging societies, there is an urgent motivation for safe, autonomous, and adaptable personal (also called *social*) robots. So humans will coexist with the next-generation robots employed as domestic workers, nurses, and caregivers at home, in hospitals, and in nursing homes.

This widespread distribution of robots will raise several completely new ethical, legal, and social issues. Robots will have the ability to learn and process our personal profiles, tastes, and habits, which will lead to privacy and safety issues, as well as those regarding individual freedom. The human–robot interactions can cause psychological and social problems, especially in vulnerable populations such as children, older persons, and patients. Then there will be issues regarding the attribution of civil and criminal liability should an autonomous robot produce damages. Finally, there will be important, critical areas bordering with bioethics, in cases of medical and biorobotics, and with humanitarian and international law, in cases of military robotics. All these cases have never been faced squarely by humanity, and this entails a need for a complex, joint approach from various disciplines to handle them.

These issues have been subject to discussion since the dawn of robotics in the works of Norbert Wiener or in the science-fiction speculations of Isaac Asimov. However, it is only in the last few years that the debate has been progressively organized within the international robotics community and that the key word *roboethics* has established itself as an emerging field of applied ethics. The complexity of the matter is enormous, as is the tableau painted by the various overlapping scientific and cultural backgrounds in the debate. This is why we believe it is worth addressing the terminology issue in this introduction to clarify the interconnecting levels between ethics and robotics.

The first level is represented by the adopted ethical theories, developed principally by the branch of philosophy called *ethics* or *morality*, which studies human conduct, moral assessments, and the concepts of good and evil, right and wrong, justice and injustice, and so on. In our case, a generic or fundamental ethical reflection is directly related to the particular issues that are generated by the development of robotic applications and their diffusion in the society. This is the proper concept of roboethics, meaning that applied ethics, similar to bioethics, attempts to provide answers to new questions that are generated by the progress of a specific scientific and technical field. This level updates the various views on concepts, such as dignity and integrity of the person and the fundamental rights of the individual, as well as the social, psychological, and legal aspects involved.

The unfolding and emerging scenarios made possible by robotics are fascinating and unsettling at the same time.

The second level, currently referred to as *robot ethics* or *machine ethics*, regards the code of conduct that designers implement in the artificial intelligence of robots. This means a sort of *artificial ethics* able to guarantee that autonomous robots will exhibit ethically acceptable behavior in all situations in which they interact with human beings or when their actions may have negative consequences on human beings or the environment. It is clear that the guidelines to define what is ethically acceptable and to enforce them are the product of the aforementioned field of roboethics. Robots are, in fact, machines, meaning tools that are unaware of the

Digital Object Identifier 10.1109/MRA.2010.940149
Date of publication: 14 April 2011

choices made by their human creators, which, therefore, bear the moral responsibility for the actions, good or bad, of robots.

Finally, there is a third level, which we could perhaps define as robot's ethics, because it is the ethic born from the subjective morality of a hypothetical robot that is equipped with a conscience and freedom to choose its own actions on the basis of a full comprehension of their implications and consequences. It is only in this case that robots may be deemed as moral agents and that one may refer to as involving the responsibilities or rights of robots. This, obviously, is currently speculative and beyond the purposes of this special issue.

It is, therefore, clear that roboethics is a work in progress, susceptible to further evolution as the events unroll in our technical and scientific future. We are convinced that all stakeholders in the development of robotics must take part, starting with the robotics scientists and also all members of the Society. The role of the media will be crucial to this: they will have to provide prompt and correct information on the progress of robotics and the pros and cons of its applications. An even more important role will be played by the world's school systems, which will have the task of training upcoming generations: the true players, beneficiaries or victims, of the imminent robotics invasion.

This widespread distribution of robots will raise several completely new ethical, legal, and social issues.

This special issue, being the first dedicated to the topic of roboethics and given the high number of submissions and the limited available space, gives priority to broader articles that provide cultural and philosophical direction to those approaching the subject for the first time and will publish some articles analyzing the human–robot relationship from various points of view: technical, psychological, sociological, and legal. Other sensitive topics, such as military robotics or biorobotics, will require further and deeper ethical analysis in future issues of the magazine. In the following paragraphs, we briefly discuss the content of each article.

The first article, "Socially Assistive Robotics," by Feil-Seifer and Matarić, examines the ethical issues involved in using socially assistive robots, particularly in the context of health care. They describe core ethical principles for robots that provide assistance through social interaction, and they emphasize how deception (intended or unintended), autonomy, and justice can affect the ethical applications of assistive robots.

The topic is further investigated in "Children, the Elderly, and Interactive Robots" by Sharkey and Sharkey, who examine the complex psychological implications of the relationships with robots, mainly through theoretical references to cognitive psychology. They start from a survey of the present state of the art in robot caregivers and pets and discuss the risks and benefits of the relational applications with the oldest and youngest members of Society.

In "The Ethical Landscape of Robotics," Łichocki et al. survey some of the main ethical issues pertaining to robotics that have been discussed in the literature so far. They start with the notion of responsibility ascription that arises when an autonomous system malfunctions or harms people. Then, they list various ethical issues emerging in two sets of robotic applications: service robots that peacefully interact with humans and lethal robots created to fight in the battlefields. Finally, they also provide a short overview of machine ethics.

Powers broadens the ongoing debate on machine ethics, adding an incremental strategy. In his approach, incrementalism in machine ethics becomes a practical proposal about how to simultaneously engineer and provide ethical sanction for robots. The article discusses the concrete proposals to do this and reflects in a critical manner on these matters.

A very interesting experimental approach is that described by Salvini et al. in "The Robot DustCart." The article describes DustCart, a project concerning the use of autonomous mobile robots to collect and transport rubbish bags in a small Italian town. After a report on the testing period (service provided, testing site, and so on), the authors deal with the social and legal implications of the experiment.

A further reflection on legal aspects is given in Asaro's article, "Remote-Control Crimes," which deals with the difficult international and cross-cultural aspects of roboethics. He discusses the difficulties of applying law to criminal activities that will be enabled in the future by new robotic capabilities, such as cybercrimes; *robot crimes* will be the subject of multiple governing laws, changing national rules, conflicting regulations, and disparate institutions.

Finally, in "Ethics in Advanced Robotics," Operto outlines a brief history of roboethics, whose development she has contributed to since its birth; in her article, she points out the need to uncover the philosophical assumptions underlying today's debate in ethical and social issues of robotics to facilitate the establishment of a common ground for the definition of principles and regulatory guidelines.

We hope that the readers will enjoy the articles in this special issue, are encouraged to deepen their interest in roboethics, and will actively contribute to the debate, which will become increasingly important with the growth of robotics in the society.

Robotics & Automation Magazine | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

Socially Assistive Robotics

Ethical Issues Related to Technology

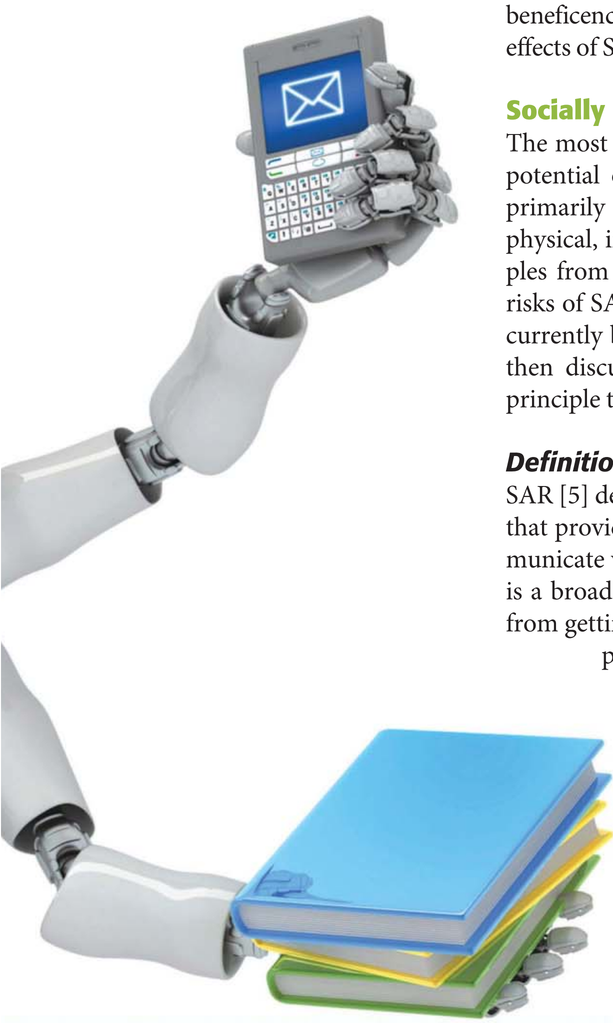
By David Feil-Seifer and Maja J. Matarić

© iSTOCK PHOTO.COM/VLADIMIR NIKITIN

Digital Object Identifier 10.1109/MRA.2010.940150
Date of publication: 14 April 2011

1070-9932/11/$26.00©2011 IEEE

Socially assistive robotics (SAR) aims to address critical areas and gaps in care by automating supervision, coaching, motivation, and companionship aspects of one-on-one interactions with individuals from various large and growing populations, including stroke survivors, the elderly and individuals with dementia, and children with autism spectrum disorders (ASDs). This article examines the ethical challenges of SAR from three points of view (user, caregiver, and peer) using core principles from medical ethics (autonomy, beneficence, nonmaleficence, and justice) to determine how intended and unintended effects of SAR can impact the delivery of care.



Socially Assistive Robotics

The most obvious and direct risk of any assistive technology, including SAR, is the potential of physical harm. While this is an important risk to examine, SAR is primarily concerned with robots that provide assistance through social, rather than physical, interaction. In this article, we outline the commonly accepted core principles from medical ethics and use those principles as guidelines for evaluating the risks of SAR. We use examples of SAR systems to describe the ways that robots are currently being used as directions for future use based on an ongoing research. We then discuss the core ethical principles to be examined. Finally, we apply each principle to SAR in turn and discuss its implications.

Definition of SAR

SAR [5] describes a class of robots that is the intersection of assistive robotics (robots that provide assistance to a user) and socially interactive robotics (robots that communicate with a user through social and nonphysical interaction). Assistive robotics is a broad class of robots whose function is to provide assistance to users, ranging from getting out of bed, brushing teeth, locomotion, and rehabilitation. This section provides few examples of SAR systems.

Wada et al. [23] describes the design of Paro, a robot for pet-therapy applications for nursing homes that do not allow pets. Pet therapy has been shown to have a positive effect on the elderly in nursing-home settings [16], but there are logistical challenges to having animals in nursing homes. Paro was built to resemble a baby harp seal and designed to interact like a pet with simple sounds and movements made in response to being held and petted. Experimental results suggest that Paro may be effective for reducing stress in nursing-home residents. In addition, when placed in common areas of nursing homes, it produced increased social activity among residents. This suggests that SAR systems may be useful not just for their direct therapeutic applications but more generally as catalysts for social interaction.

Another SAR system is Roball [18], a self-propelling robotic ball that can sense its position and motion and thus the way it is being played with. Roball is being evaluated for use by children, including children with ASDs in the home or in clinical settings. Children with ASD typically have decreased social interactive behavior; encouraging play with therapists, family, and peers and could have both diagnostic and therapeutic uses. Roball and other robots for play could be used as an addition to current ASD diagnostics or therapeutic regimens or as tools for developing new diagnostic and therapy methods. In general, the aim of SAR for ASD is to encourage children to initiate and sustain social interaction [17] with a parent, therapist, sibling, or peer.

Poststroke rehabilitation is another area where SAR can provide therapeutic benefits. Rehabilitation robotics has been developing robot arms that apply and measure forces on the user's limbs. Such hands-on movement training is particularly useful in the early stages poststroke. However, a major long-term challenge of poststroke recovery, and rehabilitation in general, is encouraging compliance with the prescribed therapeutic regimen. Matarić et al. [12] describes a SAR system designed to improve therapeutic compliance through verbal noncontact coaching and encouragement. Such systems are designed to work in concert with the established stroke exercise methods such as constraint-induced therapy, building on and augmenting effective health-care practices.

Concurrent with the developing SAR technologies, ethical appraisal studies are being conducted about their acceptance. Mutlu and Forlizzi [4] conducted an ethnographic study of a delivery robot used in multiple departments of a hospital, finding that different patient groups had different reactions to the robot. For example, cancer units were not accepting the robot, finding it annoying, while postpartum units were accepting the robot and calling it delightful. The results of this study suggest that user populations could have completely different experiences with the same robot and that these experiences could be based on the users' preexisting social and task dynamics and context. Tapus et al. [21] described a study in which elderly participants with Alzheimer's disease interacted with a SAR robot that promoted cognitive exercises through a song-recognition game in a six-month study. The study participants included the robots in their narratives and preferred it to a computer. Turkle [22] demonstrated that some participants interacting with robots can correctly identify the robot's intended emotional abilities and operational capabilities. These participants could also correctly distinguish equivalent capabilities in a person, pet, or other relational artifact. However, it was also demonstrated that some users formed attachments and emotional bonds with the robots they were interacting with. These attachments led to misconceptions about the robots' emotional capabilities. For example, one user felt that the robot would miss him when he was gone, which is something that the robot was not capable of doing. In their hyperbolic yet poignant article, Sharkey and Sharkey [19] argue that such attachments in children could lead to malformed development and emotional problems.

SAR is designed for use in a wide variety of settings including hospitals, schools, elder-care facilities, and private homes.

Persons Affected by SAR

SAR is designed for use in a wide variety of settings including hospitals, schools, elder-care facilities, and private homes. The intended end users of such systems are individuals with special needs, but SAR systems must operate in real-world environments that may also include family, caregivers, and medical personnel. Consequently, the effects of SAR must be assessed for all of the individuals affected by the technology.

Core Ethical Principles

There are many ways to approach potential ethical issues related to technology in general, and SAR in particular. Several appraisals of specific SAR systems have been implemented and some have discussed the ethical dilemmas that a particular system poses [19], [22]. Studies have also aimed to establish ethical benchmarks related to the design, manufacture, or use of SAR [7], [9]. Finally, some appraisals have applied the core ethical principles to identify potential problems [3]. In this work, we apply an established medical ethics framework to identify potential issues related to SAR. This framework uses the following core principles for considering ethical issues [4]:

- *beneficence*: caregivers should act in the best interest of the patient
- *nonmaleficence*: the doctrine, "first, do no harm," followed by the caregivers to avoid harming patients
- *autonomy*: the capacity to make an informed, uncoerced decision about care
- *justice:* fair distribution of scarce health resources.

There is dissension about whether or not the Beauchamp and Childress model is the ideal model for assessing medical ethics, as the foundation for current ethical appraisal and ethical training, we feel it is a sufficient jump-off point for discussion.

These principles underlie the ethical reviews of experiments with human participants and can also thus provide broad categories for examining ethical issues related to SAR. To perform such an examination, we use examples from actual SAR system experiments. However, these descriptions are only considerations of hypothetical scenarios and not meant to make judgments on the ethical validity of those specific SAR systems. In the next section, we describe the principles of beneficence and nonmaleficence and how they relate to the ethical use of SAR.

Beneficence and Nonmaleficence

The principles of beneficence and nonmaleficence state that caregivers should act in the best interests of the patient and should do nothing rather than take any action that may harm a patient. These principles establish that the potential benefits of an ethical treatment should exceed the risks. SAR, like any technology, features some risks along with the compelling potential benefits.

As noted earlier, SAR technologies are typically noncontact, so physical risk, while usually the most obvious ethical

Robotics & Automation MAGAZINE | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

concern, is not a major issue of concern. SAR systems are designed so the robot does not apply any forces on the user. On the other hand, the user can touch the SAR system, and in some cases (as with Paro, see earlier), such contact is part of the therapy. However, in a majority of systems no physical contact is involved, and the robot may not even be within reach of the user, though it is typically within the social interactive space conducive to one-on-one interaction through speech, gesture, and body movement.

In this section, we examine some of the aspects of SAR technologies that are unique and ways in which SAR systems, in particular, might impact not only the user directly but also others in the shared context. In particular, the most prominent nonphysical risks posed by SAR systems include, but are not limited to, attachment to the robot, deception about the abilities of the robot, and influence on the human–human interaction of a robot's user.

Relationships, Authority, and Attachment

It is safe to assume that a robot would not be the only caregiver/therapist for an assisted individual. Typically, care is provided by human caregivers, including professionals and family members. Thus, the SAR system impacts all of these individuals in various ways. For example, a robot that does something that a human caregiver would otherwise do (e.g., providing encouragement for performing exercises) might have as much impact on the human caregiver as on the patient, through the reduction of tasks related to a patient or through the reduction of workplace monotony. Specifically, many SAR systems are being designed to reduce the burden and burnout of family members and other caregivers. A SAR system might also provide a benefit to a caregiver by monitoring multiple aspects of the patient and providing ongoing quantitative assessments.

Sharkey and Sharkey [19] described another significant ethical dilemma that occurs when a user becomes emotionally attached to the robot. While establishing engagement and having the user enjoy interactions with the robot is a goal of SAR, attachment can also result in problems under certain circumstances. For example, if the robot's effectiveness wanes, its scheduled course of therapy concludes, or, if it suffers from a hardware or software malfunction, it may be taken away from the user. The robot's absence may, in cases of attachment, cause user distress and possibly result in a loss of therapeutic benefits. Attachment issues can happen with users of all ages, from children to adults and to the elderly. Such issues can be particularly acute in users who cannot understand the causes for the robot's removal but can arise even with users who have full understanding of the circumstances. Our experiments with SAR robots interacting with elderly users and users with Alzheimer's disease, mentioned earlier, demonstrate that such users do engage with robots and miss them when the robots are removed [21].

Perception and Personification of the Robot

As discussed earlier, one goal of an effective SAR system is to establish a relationship with the user that leads toward intended therapeutic goals. However, since the user cannot be fully informed about the limitations of the robot, the following issue arises: Is there deception inherent in the personification of a robot by a user or a caregiver? Such personification could be unintentional, arising from the caregiver referring to the robot as him or her, ascribing feelings to the robot, and assigning the robot greater intelligence than it may have. Studies have shown that people quickly form mental models of robots they are presented with, much as they do of people. Those models are often incorrect as they are based on what people know best: other people. The designers of the robot may purposefully manipulate the perceptions of the user toward therapeutic goals or may not intend to do so at all; in any case, if such perceptions are incorrect, the user is deceived.

Deception is a risk created by the use of robots in assistive settings. Some roles of SAR systems are most closely associated with people, such as those of a therapist, companion, teacher, or coach. In those roles, the robot may be constructed to physically resemble and act like a human equivalent. In other scenarios, the robot may fill the role of a pet or toy, with physical form to match. While it may be assumed that the physical form of the robot is deliberately designed to evoke the desired type of relationship with the user, there can be unintended ways in which the robot is perceived and received by the user. Studies of the so-called uncanny valley already demonstrate that the level of humanlike realism of the robot has an unexpected impact on people [10]. Similarly, the size of the robot has an impact on the interaction and perceived role: studies have shown that robots that approach the height and size of the user are received with some trepidation compared to smaller embodiments [11]. The way the robot is dressed and accessorized can also influence how it is perceived; a robot in a lab coat and wearing a stethoscope might be perceived as being medically competent even if it is not.

> **The most prominent nonphysical risks posed by SAR systems include attachment to the robot, deception about the abilities of the robot, and influence on the human–human interaction of a robot's user.**

The issues of physical appearance are in many ways just the tip of an iceberg; communication is also crucial. Whether the robot speaks, and if it does so, with a synthetic or recorded voice, male or female, accented or not, and containing emotion or not are all important parameters defining the nature of the interaction. These communication parameters play key roles in how effective the robot will be in a SAR setting. In addition to speech and language, embodied expression consisting of gesture, body language, and facial expressions comprises another complex area of study in human–machine interaction. This myriad of SAR design parameters has important consequences on the role of the robot and the resulting human–robot interaction; there is much research to be done in defining how factors affect interaction in general and user care in particular.

The relationship between the user and the robot as defined by the role of the robot can lead to deception with regard to the robot, contributing to increased risks. For example, a user perceiving the robot as a doctor or nurse could lead to deception. Such deception could be harmful, especially considering that the robot's communication and decision-making abilities are not on par with a human caregiver. A user could also believe that a robot is capable of assisting him/her in ways that a human would when in fact it could not. For example, if a user perceives a robot as having the abilities of a doctor, that user could equate telling the robot a pertinent piece of medical information without communicating that information to the doctor, potentially resulting in lost information. Conversely, if the user does not perceive the robot as a knowledgeable authority, he or she may not accept suggestions or instructions from the robot, thereby subverting the therapy process and rendering the robot ineffective. To complicate the matter further, such loss of authority may not be instantaneous; the user may be amenable to working with the robot for some time, perhaps due to the robot's novelty to the user, but may later lose interest in the robot.

One goal of an effective SAR system is to establish a relationship with the user that leads toward intended therapeutic goals.

Another related aspect of user perception of the robot's abilities and authority is the issue of recognition and reporting of suspect behavior. Consider a situation wherein a user is in obvious distress. A human observer would, or should, know to report the situation to an authority capable of helping. A robot, however, may not have the ability to recognize alarming behavior, yet people around the robot may believe it does and so may fail to act in response, assuming that the robot would/could handle the situation. Before any technology is deployed in an assistive setting, it is critical to establish to all involved what the capabilities of the technology are. However, as SAR and other technologies become more pervasive, uninformed bystanders will be exposed to them, and assumptions of full disclosure will quickly become unrealistic. In general, the issue of projected authority and role of the robot based on its appearance and behavior is complex, and one that could be the topic of study from a range of fields including ethics, social science, and engineering.

Changes to Human–Human Interaction

The work of Wada et al. [23] demonstrates that SAR systems can result in increased amounts of human–human interaction. However, a robot could just as easily be an isolating factor [19], [22]. Most current examples of SAR use a robot as an enhancement of the roles of current caregivers, not as their replacement, and as an addition to existing therapy, not its substitute. However, if the robot is used as a replacement or substitute for human care, then the robot might serve to reduce the amount of human–human contact. This is especially a concern if the robot is the only therapeutic influence in a user's life. For populations that are known to suffer from isolation, including the elderly or children with developmental disorders, robots might facilitate further isolation even while delivering a therapeutic benefit. We have argued that such use of technology as proxies for human attention is a real risk but not one that is new or specific to robotics. Television watching and playing computer games are both poor substitutes for attentive parenting but neither the TV nor the games can be blamed. Similarly, ethical and productive use of SAR technologies will necessarily put the burden on the caregivers to not abuse the technology.

Discussion

The core principles of beneficence and nonmaleficence are crucial for deciding whether the use of SAR is ethical and beneficial for a particular user. While there are many stated benefits for SAR in terms of encouraging social interaction and therapeutic compliance, providing therapeutic intervention and advice, there are potential ethical pitfalls. Properly describing the capabilities and role of an assistive robot is critical for caregivers to assess the potential for harm. In addition, proper communication between caregivers and users of SAR is crucial to minimize unintended deception. Finally, when robots are first introduced to users, the possibilities for upgrades or modifications that would change the robot's appearance or behavior, and the fact that the robot might or will eventually be taken away, should be made clear to the user.

Generally speaking, deception should be minimized wherever possible to avoid harm to the user. But, as noted earlier, since human perception of any part of a robot (facial expression, voice, gesture, appearance, size, etc.) is not yet well understood, unintentional interpretation and possible deception are inevitable until our understanding of the human–machine interaction is thoroughly studied and characterized. It is thus critical to conduct detailed studies in realistic but monitored settings before commercializing these technologies to improve both the safety and effectiveness of the designs.

Robotics & Automation Magazine | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

Ethical ramifications of SAR are not limited to the balance between risks and benefits. SAR also poses challenges for the user's informed decision-making ability, as discussed in the next section.

Autonomy

The core medical ethics principle of autonomy dictates that patients should be able to make informed decisions about their own care. Extending this principle to SAR, patients should be able to make informed decisions about SAR that are part of their care. As discussed in the previous section, several factors make it likely that a user may not be capable of being fully informed about the abilities and limitations of a particular SAR technology and be aware of his or her own possibly biased perceptions of it. People might believe (or be made to believe) that the robot is more capable than it is, which can create barriers to making an informed decision about care. There are also valid concerns about a user's privacy with SAR as with most other technologies. If a robot is not able to properly distinguish between confidential information (e.g., personal health information) and information that the user permits for release, then the robot may create an unintended violation of a user's privacy. In this section, we examine the problems relating to informed consent and privacy that have ethical implications. Since autonomy can also refer to robots that are in control of their own actions, we refer to patient/user autonomy as autonomy while referring to the self-control of a robot as robot autonomy or autonomous robots.

To provide the user with enough information to make an informed decision about a robot, a critical question is: Are the capabilities of an assistive robot being correctly described? If a description of how the robot will be used does not give the user the necessary information to make an informed decision about using the robot, then the caregiver is not behaving in an ethical manner. Consider the example of a companion robot for use in a nursing home that does not allow pets. If the user is told that the robot is just like a pet, but later discovers that in fact the robot only has a limited and small repertoire of behaviors, the user may become disappointed and feel lonely. However, this is not a simple issue; the robot vacuum cleaner, Roomba, is capable of very few actions related to floor vacuuming, yet studies have shown that the users of Roomba are attached to it and demand that it be fixed and returned when broken rather than that it be replaced with a new one [20]. Different users have different expectations, and so it is not necessarily possible to warn a user completely about his/her perceptions and bonding with the robot, positive or otherwise.

Similarly, the role of the robot and possible misconceptions about that role, described in the previous section, could lead a user to expect high-level humanlike medical care from a robot. While the capabilities of the robot may be effective in a specific application domain, they are not comparable to a human doctor or nurse, who may be able to assist the user with decisions or consultations outside of the prescribed therapy. If a user is anticipating an inappropriate benefit for the cost of a robot that she/he is considering purchasing, then that user is not fully informed. The impact of the decision is even more important if the user is considering an application that uses a robot in place of, rather than in addition to, a human caregiver.

The authority of the robot is another sensitive issue for SAR. A robot's intended role as a therapist may exert influence on the user, putting in question who is in control of the situation and interaction. The question, "Who is in charge?" must be addressed carefully, because the technology may require a level of authority to be effective. A user that is feeling stressed or is in pain must feel free to stop an exercise, for example, even if that is counter to the robot's advice. However, a SAR system's role in many contexts is to give direction to a user, requiring some measure of authority derived from expertise. A lack of balance between user autonomy and robot authority could create an ethical dilemma.

When discussing authority with respect to SAR, privacy is of utmost importance. A robot might not have sufficient capabilities to distinguish between privileged information and information that can be distributed. A robot may also lack the ability to distinguish between individuals who have the authority to receive information about the user and those who do not. Patients seeking medical care have an expectation of privacy backed by legal protection. However, a robot might not be able to meet these privacy obligations. In particular, a user might not realize that a robot's camera could record video, display video in another location, or that wireless transmission of video data cannot be guaranteed to be completely private. People perceive a robot's camera as having similar capabilities to human vision; this is a natural but false assumption. As discussed in the previous section, the robot might not know to communicate information that is critical to care or how to communicate privileged information in a discreet manner. Therefore, it is important to make sure that the capabilities of a robot are sufficiently explained so that a user has been well informed of a model of the robot's abilities as possible.

Television watching and playing computer games are both poor substitutes for attentive parenting but neither the TV nor the games can be blamed.

The use of SAR can also have a positive effect on the user's autonomy. An example from an assistive technology study describes how elders in independent-living situations were asked to allow cameras into their homes to allow for home monitoring for safety. The elders were uncomfortable with this process, as they did not want to be seen, especially in private places like the bathroom. The experimenters responded by using computer vision to monitor only the user's silhouettes [2], thereby providing sufficient information for the home monitoring task, but also allowing the users the autonomy in choosing what information they wished to release. SAR could employ similar techniques for allowing users privacy, thereby increasing user autonomy.

Discussion

Preserving the autonomy of a person seeking care is a core ethical value. For the most part, the procedures for informed consent are sufficient for allowing a user's autonomy in decision making regarding care. However, the potential for user deception can interfere with a user's informed consent. Currently, the appearance of a robot and its ability to sense its environment and communicate with others might not match. This mismatch might result in (unintentional) deception of the user as to the robot's capabilities, which in turn may affect the user's ability to give informed consent. To mitigate this, the users should be presented with a clear description of the robot's capabilities as well as limitations, but they must also understand that their perceptions of the robot, responses to it, and the attachments and relationships they form with it are not fully predictable, just as they are not in human–human interactions.

The determination of responsibility for a SAR's actions is a complex problem that must be addressed, as the technology is being developed and deployed.

Justice

The principle of justice governs the fair distribution of scarce resources. This can be a very difficult topic when discussing experimental treatments such as SAR. The authors know of no SAR systems that are currently used outside the research setting, so discussion of the actual cases in the field is premature. However, we can presume that for the foreseeable future, robots will be somewhat expensive. Thus, a question that should be asked is: Do the benefits of SAR outweigh the costs? Like other proposed therapies, quality of life surveys or other methods for assessing medical economy can be used to assess relative benefits, and costs can be weighed against improvements observed [1], [24]. There does not seem to be a significant difference between calculating the costs and benefits of robots compared to other assistive devices.

Another justice-related issue when discussing robotics in socially assistive settings is the notion of responsibility: Who is responsible when things go wrong? While this might not traditionally pertain to the principle of justice, fair allocation of responsibility for SAR systems might be related to a fair allocation of therapeutic resources. When a robot does not behave as intended, it could be the result of user error or it could be the result of robot error. The difference is not always readily discernable. In the case of robot error, the problem could be in the design, hardware, or software of the robot, meaning that the responsibility belongs to the designer, programmer, manufacturer, distributor, or retailer. Furthermore, the user error may be due not just to a user's self-imposed mistake but could be a result of poor training, erroneous instructions, or false expectations due to intentional deception.

Software responsibility is troubling since most software licenses explicitly absolve the software developer of responsibility. A large percentage of open-source public domain software and end-user license agreements (EULAs) specify that the software is provided as is and with no liability assumed by the developers or software companies. This includes loss of privacy or data. As privacy is a critical component of the autonomy and nonmaleficence aspects of medical ethics, such a declaration of nonresponsibility is especially concerning. It is entirely possible that a software error could leak privileged information in some way and that the software developer would feel completely within his/her rights to abdicate responsibility for such an error. From the developer's perspective, software is take-it-or-leave-it. Additionally, a developer cannot be responsible for unforeseen consequences of every line of code, especially given that hardware updates, user error, interface and power issues, and other influences can trigger software errors. This makes the notion of responsibility extremely difficult, making the enforcement of justice related to SAR a challenging prospect, considering that software is just one of the aspects of a complete SAR system.

Discussion

Challenges to the core ethical principle of justice may be the most difficult to anticipate. In fact, most of the problems associated with SAR will be discovered as the robots are used in their target domains. Currently, when robots are tested in research settings with human participants, their use, distribution, and responsibility for errors are all determined by institutional standards, and in the case of many nations, institutional review boards (IRBs). These institutions demand that the inclusion/exclusion criteria, operation of the robot, and responsibility for the robot's actions be stated in advance. Breaches of such agreements must be addressed on an individual basis, with the termination of a study as a possible consequence. However, as robots are deployed in the consumer realm, similar agreements might not be pursued.

The determination of responsibility for a SAR's actions is a complex problem that must be addressed, as the technology is being developed and deployed. It is unreasonable to assume that robots will work perfectly or be used always in a completely just and honest manner. Thus, when breakdowns occur, responsibility and restitution for any harm to a user must be assessed.

Summary

In this article, we have taken the core ethical principles from medicine as a foundation for discussing ethical issues implied by the SAR technologies being developed. Since this ethical framework was constructed with the ethical policies from the United States in mind, and the examples in this article are from North America and Japan, it is possible that different or additional ethical challenges arise for other cultures. More exploration is needed, especially to determine whether robots designed and tested in one medical care system would behave ethically in another. Additionally, as users' reactions to robots might be different from one group to the next, proven ethical principles for one user population might not be effective for another.

New technologies bring about entirely unprecedented contexts for human–machine interaction and call for thoughtful and well-informed multidisciplinary studies that include inputs and expertise and address concerns from the entire complex constituency, including the technology developers, social scientists, ethicists, and, most importantly, members of the broad user community. This process must be open and ongoing since the technologies and user responses and experiences will continue to evolve indefinitely.

References

[1] N. K. Aaronson, C. Acquadro, J. Alonso, G. Apolone, G. Apolone, D. Bucquet, M. Bullinger, K. Bungay, S. Fukuhara, B. Gandek, S. Keller, D. Razav, R. Sanson-Fisher, M. Sullivan, S. Wood-Dauphinee, A. Wagner, and J. E. Ware, Jr., "International quality of life assessment (IQOLA) project," *Quality Life Res.*, vol. 1, no. 5, pp. 349–351, Dec. 2004.

[2] D. Anderson, J. Keller, M. Skubic, X. Chen, and Z. He, "Recognizing falls from silhouettes," in *Proc. IEEE Int. Conf. Engineering in Medicine and Biology Society*. New York, 2006, pp. 6388–6391.

[3] R. Arkin, "Governing lethal behavior: Embedding ethics in a hybrid deliberative/reactive robot architecture," in *Proc. Int. Conf. Human–Robot Interaction (HRI)*, Amsterdam: The Netherlands, Mar. 2008, pp. 121–128.

[4] T. Beauchamp and J. Childress, *Principles of Biomedical Ethics*. New York: Oxford Univ. Press, 2001.

[5] D. Feil-Seifer and M. Matarić, "Defining socially assistive robotics," in *Proc. Int. Conf. Rehabilitation Robotics (ICORR)*, Chicago, IL, July 2005, pp. 465–468.

[7] D. J. Feil-Seifer, M. K. Skinner, and J. M. Matarić, "Benchmarks for evaluating socially assistive robotics," *Interact. Stud. Psychol. Benchmarks Human-Robot Interact.*, vol. 8, no. 3, pp. 423–439, 2007.

[9] H. P. Kahn, H. Ishiguro, B. Friedman, and T. Kanda, "What is a human?—Toward psychological benchmarks in the field of human–robot interaction," in *Proc. IEEE Int. Workshop Robot and Human Interactive Communication (RO-MAN)*, Hatfield, U.K., Sept. 2006, pp. 364–371.

[10] T. Kanda, H. Ishiguro, M. Imai, and T. Ono, "Development and evaluation of interactive humanoid robots," *Proc. IEEE (Special Issue on Human Interactive Robot for Psychological Enrichment)*, vol. 92, pp. 1839–1850, Nov. 2004.

[11] M. K. Lee, J. Forlizzi, P. E. Rybski, F. Crabbe, W. Chung, J. Finkle, E. Glaser, and S. Kiesler, "The Snackbot: Documenting the design of a robot for long-term human–robot interaction," in *Proc. Int. Conf. Human Robot Interaction (HRI)*. New York, NY: ACM, pp. 7–14.

[12] M. Matarić, J. Eriksson, D. Feil-Seifer, and C. Winstein, "Socially assistive robotics for post-stroke rehabilitation," *J. Neuroeng. Rehabil.*, vol. 4, no. 5, Feb. 2007.

[14] B. Mutlu and J. Forlizzi, "Robots in organizations: The role of workflow, social, and environmental factors in human-robot interaction," in *Proc. Int. Conf. Human Robot Interaction (HRI)*, Amsterdam, The Netherlands. New York: ACM, Mar. 2008, pp. 287–294.

[16] I. Perelle and D. Granville, "Assessment of the effectiveness of a pet facilitated therapy program in a nursing home setting," *Society Animals*, vol. 1, no. 1, pp. 91–100, 1993.

[17] B. Robins, N. Otero, E. Ferrari, and K. Dautenhahn, "Eliciting requirements for a robotic toy for children with autism-results from user panels," in *Proc. IEEE Int. Symp. Robot and Human Interactive Communication*, Jeju Island, South Korea, 2007, pp. 101–106.

[18] T. Salter, F. Michaud, D. Létourneau, D. Lee, and I. Werry, "Using proprioceptive sensors for categorizing interactions," in *Proc. Human–Robot Interaction (HRI)*, Washington, DC, Mar. 2007, pp. 105–112.

[19] N. Sharkey and A. Sharkey, "The crying shame of robot nannies: An ethical appraisal," *Interact. Stud.: Social Behav. Commun. Biol. Artif. Syst.*, vol. 11, no. 2, pp. 161–190, 2010.

[20] J. Sung, L. Guo, R. Grinter, and H. Christensen, "My Roomba is Rambo: Intimate home appliances," in *Proc. UbiComp 2007: Ubiquitous Computing*, pp. 145–162.

[21] A. Tapus, C. Tapus, and M. Matarić, "The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia," in *Proc. Int. Conf. Rehabilitation Robotics (ICORR-09)*, Kyoto, Japan, June 23–26, 2009, pp. 924–929.

[22] S. Turkle, "Relational artifacts/children/elders: The complexities of cybercompanions," in *Proc. COGSCI 2005 Workshop Toward Social Mechanisms of Android Science*. Stresa, Italy, July 2005, pp. 62–73.

[23] K. Wada, T. Shibata, T. Saito, K. Sakamoto, and K. Tanie, "Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, Taipei, Taiwan, Sept. 2003, pp. 2785–2790.

[24] S. Wood-Dauphinee, "Assessing quality of life in clinical research: From where have we come and where are we going?" *J. Clin. Epidemiol.*, vol. 52, no. 4, pp. 355–363, Apr. 1999.

David Feil-Seifer, Interaction Laboratory, Center for Robotics and Embedded Systems, Department of Computer Science, University of Southern California, Los Angeles, CA, 90089-0781, USA. E-mail: dfseifer@usc.edu.

Maja J. Matarić, USC Center for Robotics and Embedded Systems, Robotics Research Lab, and Viterbi School of Engineering, University of Southern California, Los Angeles, CA.

© DIGITAL STOCK & PHOTODISC

It is becoming easier to make robots that seem to understand us and even appear to be like us. There are humanoid robots that can walk, talk, and even shake your hand. There are robots that can recognize human emotional expressions and display emotional signals. There are robots that can recognize particular individuals. There are robot pets that respond to affection and that seem to need looking after. Considerable efforts are being directed toward the development of robots that people enjoy interacting with and want to spend time with. At the same time, developments in robotics are reaching the point where robot caregivers and companions for vulnerable members of society are becoming a real possibility [1], [2]. Before progressing too far down the road toward robot care, it is important to consider what ethical problems are involved in allowing, or even encouraging, the youngest and the eldest members of the society to think that they can form relationships with robots.

The idea of developing robot companions and caregivers for the elderly is taking hold. Elderly people are often lonely and in need of companionship and social contact. Some hold that a robot could be a friend substitute and, at the same time, reassure absent families about the well-being of their elderly relative by monitoring and reporting on their health. Alzheimer's disease leaves many elderly confused so that they need help with routine activities and someone to answer their questions. It has been suggested that a robot could fulfill this role. Young children need constant care and supervision, but busy

Children, the Elderly, and Interactive Robots

Anthropomorphism and Deception in Robot Care and Companionship

By Amanda Sharkey and Noel Sharkey

Digital Object Identifier 10.1109/MRA.2010.940151
Date of publication: 14 April 2011

parents do not always have the time to provide it. What harm would there be in a robot nanny taking over some of the care?

Robot companions for the very old and robot nannies for the very young are likely to be designed so that their appearance, movements, and interactions foster the attribution of mental states to them. The aim would be to provide the robots with sufficient features to encourage the target groups to form a relationship with them. Is this a form of deception and is it ethically acceptable? Our focus here is on the ethical issues involved in creating and promoting the illusion of animacy for care robots. We ask what are the pros and cons of encouraging anthropomorphic beliefs in either the elderly or the very young?

Robot Caregivers and Robot Pets

The likely development of robot caregivers for the elderly is illustrated by Gecko Systems International Corp.'s predictions that its sales of eldercare personal robots will reach US$8.3 billion by 2014. They are developing the CareBot, a personal robot equipped with multiple vital sign sensors that can follow an elderly person in their home: home-evaluation trials with the elderly began in November 2009. They suggest that the CareBot could become "a new kind of companion that always stays close to them enabling friends and family to care from afar." The CareBot is capable of verbal interaction, the delivery of medicine, video monitoring, and two-way interactions. Robotsoft's Kompai robot is similarly proposed as an aide for the elderly. It too can speak and respond to voice commands. Currently, it has an unchanging face, but there are plans to give it the ability to make emotional expressions [3].

Robot caregivers could also be used to look after children, as the Gecko Web site for the CareBot suggests. Other robots have also been developed with childcare in mind. The childcare version of PaPeRo enables mobile monitoring of children. Cameras in the robot's eyes can transmit images of the child to a window on the parent-care giver's computer or to their mobile phone. The care giver can see and control the robot to find the child if she moves out of sight. The Hello Kitty robot has a moving head and arms. Despite its limited mobility, the Hello Kitty robot was marketed on some sites as a robot child care giver: "This is a perfect robot for whoever does not have a lot time [*sic*] to stay with their child" [2]. [A more recent version (14 September 2010) says, "This is a perfect robot for times when your child needs a little extra comfort and friendship. This Hello Kitty robot will keep your child happily occupied" (http://www.dreamkitty.com).

In addition to robots that are developed with the aim of supervision and monitoring, there has been a considerable interest in the development of robot pets to act as companions. These include Paro, a fur-covered robotic seal, which was specifically designed for therapeutic uses with the elderly. Developed by National Institute of Advanced Industrial Science and Technology (AIST), it responds to petting by moving its tail and opening and closing its eyes. It reacts to sounds and can learn to respond to its name. It makes seallike sounds and is active in the day, preferring to sleep at night. It can detect light and dark by means of a light sensor and recognize when it is being held, stroked, or hit by means of posture and tactile sensors. Sony's artificial intelligence robotic (AIBO) dog, developed as an entertainment robot, has also been used in robot companions research. It has a metallic doglike form and can walk or chase a ball. It has sensors that can detect distance, acceleration, sound, vibration, and pressure. It can express six emotions (happiness, anger, fear, sadness, surprise, and dislike) by means of its tail, body movements, and the color and shape of its eyes. More recent versions can recognize voice commands, and the robot slightly exhibits different behavior depending on the interactions it has experienced.

Other artifacts have been touted as possible companions for the elderly [4]. Toy robots that could entertain the elderly (or children) include: Pleo, Ifbot, and Primo Puel. Pleo is a robotic dinosaur with many sensors that respond with different behaviors depending on its treatment. Ifbot was developed by Business Design Laboratory Co. for elderly people and can converse with them by means of a large number of stored interaction patterns. Primo Puel is an interactive doll that talks, giggles, and asks for cuddles. It was originally designed to stand in for a boyfriend for young single women but proved unexpectedly popular with elderly women in Japan.

Anthropomorphism and Deception

Although some of the robots described earlier have practical purposes, e.g., medical monitoring, most of them have features that persuade people to interact with them and form seeming relationships with them: in other words, to encourage them to be anthropomorphic or zoomorphic toward them.

Anthropomorphism is the term used to describe the behavior of attributing humanlike properties and mental states to nonhuman agents and objects. Zoomorphism is a related concept applied to the attribution of animal characteristics to nonanimals. Robots that move in a human or animallike way (and/or those that have humanoid or animallike appearances) can encourage anthropomorphism or zoomorphism.

There is a plethora of active ongoing robotics projects aimed at increasing the believability of human–robot interaction. One such area of research is the incorporation of touch sensitivity. The PaPeRo robot has touch sensors on its head and body and can tell if it is being patted or hit. The Huggable [5] has a dense sensor network for detecting the affective component of touch in rubbing, petting, tapping, scratching, and other types of interactions that a person normally has with a pet animal. It seems clear that a robot responding contingently to touch by purring or making pleasing gestures will increase its appeal. For example, Tanaka et al. [6] reported that children were more interested in the quest for curiosity (QRIO) robot that inhabited their nursery when they discovered that patting it on the head caused it to giggle.

Spoken language is a key element in human–robot interaction. Many robots have some ability to recognize and

Robotics & Automation MAGAZINE | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

respond to speech. For example, iRobi (by Yujin Robotics of South Korea) responds to 1,000 words of voice commands. The Kompai robot responds to voice commands and speaks. The Gecko CareBot also responds to voice commands. PaPeRo recognizes about 200 words and gets out of conversational difficulties by making jokes or dancing. Current robots do not have a full-blown natural language-processing interface, yet they can often create the illusion of understanding.

Face recognition is another important factor in developing relationships. Some care robots are already able to store and recognize a limited number of faces, allowing them to distinguish between people and call them by name. An even more compelling way to create the illusion of a robot having mental states and intentions is to give it the ability to recognize the emotion conveyed by a person's facial expression. Research in emotional expression recognition has been proceeding apace: smile-detection algorithms are incorporated in many digital cameras, and a recent article reports the use of machine-learning methods to distinguish between facial expressions, indicating real or posed pain responses [7]. The development of flexible skinlike materials for robot faces also facilitates their ability to make convincing emotional expressions, as in the Albert Einstein head designed by David Hanson, and augmented with recognition software by the Machine Perception Laboratory at University of California, San Diego.

Robots can be programmed to react politely to us, imitate us, and behave acceptably in the presence of humans [8]. It is possible to make people believe that robots can understand them at least some of the time. Advances in language processing, touch, and expression recognition will act to strengthen the illusion of animacy and sentience and could strengthen human–robot relationships and maintain them for longer.

Should we see efforts to develop features that promote the illusion of mental life in robots as forms of deception? In an important sense they must be, since current robots do not have minds or experiences (in this, we ignore the ongoing debates about whether in future there will be sentient artificial intelligence programs or robots). The question then is, should attempts to create an illusion of robot sentience to foster the belief that a robot is something or someone worth forming a relationship would be viewed as both deceptive and unethical?

Some have argued that this is the case. Robert Sparrow, in particular, has suggested, in the context of a discussion of the possibility of robot pet companions, that any resulting benefits for the elderly,

> are predicated on mistaking, at a conscious or unconscious level, the robot for a real animal. For an individual to benefit significantly from ownership of a robot pet, they must systematically delude themselves regarding the real nature of their relation with the animal. It requires sentimentality of a morally deplorable sort. Indulging in such sentimentality violates a (weak) duty that we have to ourselves to apprehend the world accurately. The design and manufacture of these robots is unethical in so far as it presupposes or encourages [9].

Sparrow [9] and Sparrow and Sparrow [10] argued that any beneficial effects of robot pets or companions are a consequence of deceiving the elderly person into believing that the robot pet is something with which they could have a relationship. Wallach and Allen [11], in a discussion of the ability of robots to detect basic human social gestures and respond with humanlike social cues, suggest that, "from a puritanical perspective, all such techniques are arguably forms of deception" [11, p. 44].

Should we then conclude that all attempts to induce the illusion of sentience in machines are unethical? We suggest not. Although much of the artificial intelligence depends on creating illusions, and in that sense is a form of deception [12], such a conclusion seems too extreme. The issue of deception is not a straightforward one. It is complicated by the possible anthropomorphic contribution of the viewer. For instance, Zizek [13] describes how people can choose to act as though something were real, "I know very well that this is just an inanimate object, but nonetheless I act as if I believe that this is a living being." People are anthropomorphic about far more than robots—they often behave as though objects such as their computer or their car were alive (particularly, when things are not behaving as expected). Also, views about artifacts like robots may be unclear—they may be seen neither as being sentient nor as objects but as falling betwixt and between known categories, as discussed by Turkle et al. [14].

Children enjoy make-believe play and let's pretend games. As Cayton [15, p. 283] points out, "When children play make-believe and let's pretend games, they absolutely know it is pretend . . . Real play is a conscious activity. Ask a child who is playing with a doll what they are doing and they may tell you matter-of-factly that they are going to the shops or that the doll is sick, but they will also tell you that they are playing."

A puppet, on the other hand, is outside of the child's control and less imagination and pretence is required. But a child left alone with a puppet soon realizes the illusion. The difference with a robot is that it can still operate and act when the child is alone with it. This could create physical, social, and relational anthropomorphism that a child might perceive as real and not illusion. Young children may not know enough about technology to understand the differences between living creatures and convincing robots. The same distinction might be difficult for elderly people with Alzheimer's disease.

In addition, elements of anthropomorphism may be beyond conscious control. People might report knowing that the robot they are interacting with is a machine, but may nonetheless respond to it in some ways as if it were alive. Epley et al. [16] suggest that even metaphorical invocations of anthropomorphism may have an effect on behavior: "Metaphors that might represent a very weak form of anthropomorphism can still have a powerful impact on behavior, with people behaving toward agents in ways that are consistent with these metaphors."

It may be that a robot that seems to resemble us, or to respond to us, will inevitably be anthropomorphized to some degree. Designing robots to encourage anthropomorphic attributions could therefore be viewed as an unethical form of deception. However, in that case, giving any object a human or animallike appearance could also be seen as deception. It seems too extreme to suggest that dolls, puppets, and statues should no longer be made or played with. People, in general, and children, in particular, exhibit anthropomorphic behavior much of the time. Anthropomorphic design occurs in many more areas than robotics, from Alessi bottle openers to car grilles and even pet rocks [17]. Rather than objecting to all such uses, it makes more sense to focus our ethical concern on those situations in which anthropomorphic design seems likely to lead to negative consequences for human welfare. Some such consequences are considered in the following section.

A further cause for concern is that there are reasons to expect the vulnerable youngest and eldest members of society to be more likely to be affected by anthropomorphism. Both have a strong need for social contact, and both may lack knowledge of the technology underlying the apparent responsiveness of interactive robots. Both these factors have been argued to increase the tendency to be anthropomorphic in recent accounts [16].

Epley et al. [16] argue that the tendency to anthropomorphize nonhuman agents depends on three psychological determinants: the accessibility and applicability of anthropocentric knowledge, the motivation to explain and understand the behavior of other agents, and the desire for social contact. Their argument is backed up by extensive experimental evidence, of which a few examples are cited here. Various factors can be shown to affect the accessibility and applicability of anthropocentric knowledge: for example, greater similarity between the appearance and behavior of an entity and humans, or animals, can increase the degree of anthropomorphism and empathy shown toward it. Thus, DiSalvo et al. [18] found that robots are anthropomorphized more readily when given humanlike faces and bodies. The idea that anthropomorphism is stronger when there is a need to explain is supported by the evidence that shows that unpredictable behavior increases the tendency for anthropomorphic explanations [19]. Finally, in accord with the desire for social contact determinant, experimental manipulations show that, when feelings of loneliness are induced, people are more likely to anthropomorphize pets and gadgets [20].

This account of anthropomorphism can be used to argue that both the very young and the very old may be more likely than other age groups to be anthropomorphic and less able to understand the limited ability of robots to understand and empathize. Both groups have a strong desire for social contact: babies (because they are innately predisposed to look for human social contact) and the elderly (because they are often lonely). In addition, both are likely to lack knowledge about how robots work. Infants and young children are not clear about the differences between living and nonliving entities [21]. Elderly people with Alzheimer's may not be able to understand the mechanisms underlying robot behavior. Both groups might be more prepared to form relationships with robots and robot pets designed to give the illusion of sentience than other groups of the population.

Both the young and old may show a stronger tendency to anthropomorphize robot companions and pets, but whether or not this is amounts to an ethical problem depends in part on what the consequences of such anthropomorphism might be. We consider these in the following sections.

Likely Consequences for Robots and the Elderly

One negative consequence of an elderly person imagining that they have a relationship with a robot might be an increase in their level of anxiety—they might think that they had to look after the robot, even at the expense of their own well-being. Observers and relatives of a confused old person looking after a robot pet might see it as depriving their relative of dignity and infantilizing them.

Similar points have been made in the context of the doll therapy that has been undertaken with those with Alzheimer's disease. Positive effects have been found from doll therapy, where dolls are given to clients to stimulate memories of a rewarding life role, especially that of a parent, and to act as a focus for reminiscence and conversation [15]. However, ethical objections have been raised to the effect that doll therapy infantilizes the elderly [15].

Studies have shown that clients with dementia engaged in doll therapy tend to believe that their dolls are real babies. When Mackenzie et al. [22] questioned the care workers in homes where doll therapy had been tried, they discovered that some residents would put the doll's interests before their own as one would with a real baby. They also found that some caregivers, visiting relatives, and fellow residents saw the doll therapy as demeaning and patronizing.

Looking after robot pets could be seen to similarly infantilize elderly people, although a mitigating factor is that robots can be seen as cool gadgets in a way that dolls are not. Another possible negative consequence is that the presence of a robot might result in a reduction in the level of social interaction an elderly person experiences. An outcome in which an elderly person chose to spend time with the robot rather than taking part in social interactions with humans would be unwelcome. Similarly, if other people were to assume that the social needs of an elderly person were being taken care of by the robot and so interacted less with them, that would also be a problem.

On the other hand, there are reasons to expect some positive outcomes. Various studies have found evidence that the elderly can benefit from interacting with robot companions. The positive effects are said to be similar to those obtained from animal-assisted therapy [23]. For instance, Kanamori et al. [24] showed various improvements in elderly persons who interacted regularly with a Sony AIBO robotic dog—their loneliness scores were reduced, and their quality of life assessment scores improved. Banks et al. [25] even found that elderly people in long-term care facilities benefited as much from interacting with an AIBO robotic dog as from interacting with a real dog. Elderly dementia patients have also shown positive outcomes, including increased communication as a result of sessions with an AIBO [26].

It is of course important to be cautious about the interpretation of such studies. The positive effects depend on comparisons with a control measure. The results reported by Kanamori et al. [24] showed improvements in well-being over time between initial and later sessions. Banks et al. [25] showed that beneficial effects were obtained for those interacting with either the real or the robotic dog, when compared with the control group who received no such opportunities for interaction. However, such improvements could have been found because the alternative was so dire. Someone in solitary confinement might benefit from being given a robot companion, but they would benefit far more from a friendly social environment. It is not clear that the same relative improvements would be found if the comparison were to a control group that received other forms of intervention, such as a visit by someone who chatted and held their hand. It is also important to check that any benefits are maintained over time. An initially interesting robot may rapidly lose its appeal.

Nonetheless, the elderly might obtain some health benefits from interacting with a robot. The robots could also stimulate further social interaction with other people. Robot pets can act as social facilitators, leading to increased interactions between their elderly owners and other people. Robot toys can give an elderly person something to talk about and other people something to talk to them about. For instance, when Wada and Shibata [27] videoed interactions between a Paro robot seal and a group of elderly care home residents, they found that the social interactions between the residents themselves increased at the same time that physiological indicators showed reduced stress levels. It seems that Paro even encouraged positive communication and resulted in a reduction of the backbiting that had previously characterized their interactions.

A robot that facilitates conversation may function as an attractor for visitors. Children may want to play with the robot and have fun with granddad's big toy. Kanamori et al [24] report the case of an 84-year-old man who talked much more to his children after the introduction of an AIBO robot dog. It gave both him and them a focused object to talk about. In such cases, the underlying deceptive illusion could be justified. Nonetheless, a more utopian vision in which the frail elderly experienced real caring relationships with humans would still seem preferable to a world in which the meaning of their lives depended on animated machines.

Consequences for Babies and Children

Some positive outcomes could result from the combination of elderly people and interactive robots. Positive consequences seem less likely in the case of babies and young infants. Because these youngest members of society have a strong social drive and a lack of knowledge about technology, they are particularly likely to overestimate the abilities of robots that have some of the features of humans or animals. There is a risk that such overestimation by the infants themselves, and by those around them, might result in them spending too much time with robots. This could diminish the time they get to spend in the company of a sensitive human caregiver and impede the development of their understanding of how to interact with fellow human beings.

Infants need to form attachments to a significant caregiver. The kind of attachment they form has a strong influence on their subsequent development. It is well known that, for an infant to become well adjusted and socially attuned, they need a caregiver with sufficient maternal sensitivity to perceive and understand their cues and respond to them promptly and appropriately [28]. It is this that promotes the development of secure attachment in infants and allows them to explore their environment and develop socially. There are disturbing illustrations of the effects of being raised in the absence of human attachment figures in reports of the development of those raised in the impoverished conditions of Romanian orphanages. Nelson et al. [29] compared the cognitive development of young children reared in Romanian institutions to that of those moved to foster care with families. The results showed that children reared in institutions manifested greatly diminished intellectual performance (borderline mental retardation) compared with children reared in families. Chugani et al. [30] found that Romanian orphans who had virtually experienced no mothering differed from children of comparable ages in their brain development and had less active orbitofrontal cortex, hippocampus, amygdale, and temporal areas.

There is little reason to suppose that a robot could provide an adequate replacement for human care. As discussed by Sharkey and Sharkey [2], it is unlikely that a robot would be able to respond to a child in the sensitive manner needed to engender secure attachment. Secure attachment to a caregiver is associated with better development in part because of what the infant learns as a result.

A securely attached child learns to take another's perspective. When the mother reflects their baby's emotional distress in their facial expression, it helps the baby form a representation of their own emotions. This social biofeedback leads to the development of a second-order symbolic representation of the infant's own emotional state [31], [32] and facilitates the development of the ability to empathize and understand the emotions and intentions of others. These are not skills that any near-future robot is likely to have. What patterns of social behavior and reciprocal interaction would a baby learn from a robot that responded contingently to it?

Spending too much time in the company of a robot is unlikely to help and could interfere with an infant's learning about the give and take of human relationships. Similarly, a robot is not going to be an adequate replacement for a parent in terms of an infant's linguistic development. Advances in natural language processing could lead to superficially convincing conversations between robots and children in the near future. However, such interactions would not be meaningful in the way that caring adult–child interactions are. It is one thing for a machine to give a convincing conversational response to a remark or question and a completely different thing to provide appropriate guidance or well-founded answers to puzzling cultural questions. There are many cues that an adult human uses to understand what answer the child requires and at what level.

Language interactions between young children and adults are transactional in nature, both participants change over time. Adults change register according to the child's abilities and understanding. They continuously assess the child's comprehension abilities through both language and nonverbal cues and push along the child's understanding. This is required for both language development and cognitive development in general. It would be extremely difficult to find specifiable rules that a robot could apply for transactional communication to adequately replace a care giver's intuitions about appropriate guidance.

Babies and infants would probably not be able to resist interacting with a robot that responded to them contingently. There are, however, reasons to fear the effects of such interactions, given that an infant's experiences of interactions have such a powerful effect on their development. What an infant learns about getting a response from a robot nanny is unlikely to help it understand the subtle and nuanced reciprocal interactions that are needed to form good human relationships. It might seem convenient to have a robot entertain your baby so that you can get some more work done, but the risks might be too great.

In addition to impeded social, emotional, and linguistic development, a young child spending too much time with a robot might suffer other negative consequences. Bryson [33] points out that interactions with robots will be much more predictable than interactions with humans and that children might come to prefer this. In a related argument, Kubinyi et al. [34] argue that just as cross-fostered animals and birds learn behaviors and responses when raised by those of a different species, so humans raised by robots might develop differently. They might, for instance, grow up dependent on individualized entertaining systems and be socialized to follow nonhuman behavioral patterns. A new form of human, homo technicus might emerge [34]. Melson [35] also considers the effects of adapting to pseudointeractions with technology and suggests that if children begin to think about robots as being alive, they may also begin to think about humans and animals in more mechanistic terms and with less regard to their moral standing.

There are considerable risks of negative consequences from leaving babies and infants in the company of robots. The same is not necessarily true for older children. For children who have formed secure attachments to human caregivers and who have a good grounding in human–social interaction, some exposure to robots might even be useful. Since robots will probably play an increasingly important role in society, it would be just as well if children were educated about their workings and familiar with them. Melson [35] suggests that robotic literacy should be encouraged for both parents and children. "Such 'literacy' would help adopters of this technology understand: 1) how robots are produced, maintained, and operated, emphasizing their human-produced properties; 2) what the limits and potentials are for various robotic technologies; and 3) what the distinctions are between living and 'pretend' living—stuffed animals, puppets, and robots." In addition, encouraging children and adults to understand the nature of anthropomorphism and the methods that can be used to strengthen the illusion of mental states in nonliving machines could be a powerful way of protecting them from the ill effects that might result from overestimating the abilities of robots.

Conclusions

Clearly, there is a growing interest in developing robot caregivers and companions, particularly for the youngest and eldest members of society. At the same time, there is an ever-increasing ability to implement design features that create the illusion that robots are sentient and able to respond emotionally to us. Such developments raise the likelihood that vulnerable members of society will be left in the company of robots and that people will mistakenly believe that the robots are capable of caring for them and forming mutual relationships. In this article, we have probed the ethics of designing robots that promote the illusion of being able to form meaningful relationships with humans.

It is acknowledged that some form of deception is involved in developing robots that appear to understand us. However, this deception depends on exploiting the natural anthropomorphism of the observer. Anthropomorphic design is prevalent in many aspects of society outside of robotics, and to an extent, being anthropomorphic may be an unavoidable part of being human. Clearly, it would be unreasonable to call all such design unethical. Our arguments are focused on cases where the probable consequences are a reduction in well-being.

It is suggested that, for various reasons, the young and the elderly are likely to be particularly susceptible to such designs. We conclude that robot companions for the elderly could offer positive benefits in terms of improvements in health and welfare, although these are risks in terms of dignity and loss of social contact. In contrast, the development of robot companions and caregivers for babies and infants are more likely to lead to negative consequences. The attachments that infants form with human caregivers fashion the basis of their emotional and social development, and infants that spent too long interacting with robots could learn aberrant forms of interaction. There are reasons to be ethically concerned about the possible effects of exposing either of these vulnerable groups of society to robot care and companionship, but in the case of the very young, the dangers seem to clearly outweigh any advantages.

References

[1] N. Sharkey and A. Sharkey, "Living with robots: Ethical tradeoffs in eldercare," in *Close Engagements with Artificial Companions: Key Psychological, Social, Ethical and Design Issues*, Y. Wilks, Ed. Amsterdam: John Benjamins, 2010, pp. 245–256.

[2] N. Sharkey and A. Sharkey, "The crying shame of robot nannies: An ethical appraisal," *Interact. Stud.*, vol. 11, no. 2, pp. 161–190, 2010.

[3] B. T. Horowitz, "Cybercare: Will robots help the elderly to live at home for longer?," *Sci. Amer.*, June 21, 2010.

[4] Y. Wilks, Ed., *Close Engagements with Artificial Companions: Key Psychological, Social, Ethical and Design Issues*. Amsterdam: John Benjamins, 2010.

[5] W. D. Stiehl, J. Lieberman, C. Breazeal, L. Basel, and L. Lalla, "The design of the huggable: A therapeutic robotic companion for relational affective touch," presented at AAAI Fall Symp. Caring Machines: AI in Eldercare, Washington, DC, 2005.

[6] F. Tanaka, A. Cicourel, and J. R. Movellan, "Socialization between toddlers and robots at an early childhood education center," *Proc. Nat. Acad. Sci. USA*, vol. 194, no. 46, pp. 17954–17958, 2007.

[7] G. C. Littlewort, M. S. Bartlett, and K. Lee, "Automatic coding of facial expressions displayed during posed and genuine pain," *J. Image Vis. Comput.*, vol. 27, no. 12, pp. 1797–1803, 2009.

[8] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robot. Autonom. Syst.*, vol. 42, no. 3-4, pp. 143–166, 2003.

[9] R. Sparrow, "The march of the robot dogs," *Ethics Inform. Technol.*, vol. 4, no. 4, pp. 305–318, 2002.

[10] R. Sparrow and L. Sparrow, "In the hands of machines? The future of aged care," *Mind Mach.*, vol. 16, no. 2, pp. 141–161, 2006.

[11] W. Wallach and C. Allen, *Moral Machines: Teaching Robots Right from Wrong*. New York: Oxford Univ. Press, 2009.

[12] N. Sharkey and A. Sharkey, "Artificial intelligence and natural magic," *Artif. Intell. Rev.*, vol. 25, no. 1-2, pp. 9–19, 2006.

[13] S. Zizek, *The Zizek Reader*. London: Blackwell, 2002.

[14] S. Turkle, W. Taggart, C. D. Kidd, and O. Dasté, "Relational artifacts with children and elders: The complexities of cybercompanionship," *Connect. Sci.*, vol. 18, no. 4, pp. 347–362, 2006.

[15] H. Cayton, "From childhood to childhood? Autonomy and dependence through the ages of life," in *Dementia: Mind, Meaning, and the Person*, J. C. Hughes, S. J. Louw, and S. R. Sabat, Eds. Oxford, U.K.: Oxford Univ. Press, 2006, pp. 277–286.

[16] N. Epley, A. Waytz, and J. T. Caciopo, "On seeing human: A three factor theory of anthropomorphism," *Psychol. Rev.*, vol. 114, no. 4, pp. 864–886, 2007.

[17] T. Kiesler and S. Kiesler, "My pet rock and me: An experimental exploration of the self extension concept," *Adv. Consumer Res.*, vol. 32, pp. 365–370, 2005.

[18] C. DiSalvo, F. Gemperle, J. Forlizzi, and S. Kiesler, "All robots are not created equal: Design and the perception of humanness in robot heads," presented at the DIS2002 Conf. Proc., London, England, June 2002.

[19] A. Waytz, C. Morewedge, N. Epley, G. Monteleone, J. Gao, and J. T. Cacioppo, "Making sense by making sentient: Effectance motivation increases anthropomorphism," *J. Personal. Social Psychol.*, vol. 3, no. 99, pp. 410–435, 2010.

[20] N. Epley, S. Akalis, A. Waytz, and J. T. Cacioppo, "Creating social connection through inferential reproduction: Loneliness and perceived agency in gadgets, gods, and greyhounds," *Psychol. Sci.*, vol. 19, no. 2, pp. 114–120, 2008.

[21] D. H. Rakison and D. Poulin-Dubois, "The developmental origin of the animate-inanimate distinction," *Psychol. Bull.*, vol. 127, no. 2, pp. 209–228, 2001.

[22] L. Mackenzie, I. James, R. Morse, E. Mukaetova-Ladinska, and F. K. Reichelt, "A pilot study on the use of dolls for people with dementia," *Age Ageing*, vol. 35, no. 4, pp. 441–444, 2006.

[23] M. R. Banks and W. A. Banks, "The effects of animal-assisted therapy on loneliness in an elderly population in long-term care facilities," *J. Gerontol. A Biol. Sci. Med. Sci.*, vol. 57, no. 7, pp. M428–M432, 2002.

[24] M. Kanamori, M. Suzuki, and M. Tanaka, "Maintenance and improvement of quality of life among elderly patients using a pet-type robot," *Japan. J. Geriat.*, vol. 39, no. 2, pp. 214–218, 2002.

[25] M. R. Banks, L. M. Willoughby, and W. A. Banks, "Animal-assisted therapy and loneliness in nursing homes: Use of robotic versus living dogs," *J. Amer. Med. Dir. Assoc.*, vol. 9, no. 3, pp. 173–177, 2008.

[26] T. Tamura, S. Yonemitsu, A. Itoh, D. Oikawa, A. Kawakami, Y. Higashi, T. Fujimooto, and L. Nakajima, "Is an entertainment robot useful in the care of elderly people with severe dementia?," *J. Gerontol. A Biol. Sci. Med. Sci.*, vol. 59, no. 1, pp. 83–85, 2004.

[27] K. Wada and T. Shibata, "Robot therapy in a care house: Its sociopsychological and physiological effects on the residents," in *Proc. 2006 Int. Conf. Robotics and Automation*, Orlando, FL, May 2006, pp. 3966–3971.

[28] M. D. S. Ainsworth, S. M. Bell, and D. J. Stayton, "Infant–mother attachment and social development: Socialisation as a product of reciprocal responsiveness to signals," in *The Introduction of the Child into a Social World*, M. P. M. Richards, Ed. London: Cambridge Univ. Press, 1974, pp. 9–135.

[29] C. A. Nelson, C. H. Zeanah, N. A. Fox, P. J. Marshall, A. T. Smyke, and D. Guthrie, "Cognitive recovery in socially deprived young children: The Bucharest early intervention project," *Science*, vol. 318, no. 5858, pp. 1937–1940, 2007.

[30] H. Chugani, M. Behen, O. Muzik, C. Juhasz, F. Nagy, and D. Chugani, "Local brain functional activity following early deprivation: A study of post-institutionalised Romanian orphans," *Neuroimage*, vol. 14, no. 6, pp. 1290–1301, 2001.

[31] P. Fonagy, "The development of psychopathology from infancy to adulthood: The mysterious unfolding of disturbance in time," *Infant Mental Health J.*, vol. 24, no. 3, pp. 212–239, May/June 2003.

[32] G. Gergely and J. Watson, "Early social-emotional development: Contingency perception and the social biofeedback model," in *Early Social Cognition: Understanding Others in the First Months of Life*, P. Rochat, Ed. Hillsdale, NJ, Erlbaum, 1999, pp. 101–137.

[33] J. Bryson, "Why robot nannies probably won't do much psychological damage," *Interact. Stud.*, vol. 11, no. 2, pp. 196–200, 2010.

[34] E. Kubinyi, P. Pongracz, and A. Miklosi, "Can you kill a robot nanny? Ethological approach to the effect of robot caregivers on child development and human evolution," *Interact. Stud.*, vol. 11, no. 2, pp. 214–219, 2010.

[35] G. F. Melson, "Child development robots: Social forces, children's perspectives," *Interact. Stud.*, vol. 11, no. 2, pp. 227–232, 2010.

Amanda Sharkey, Department of Computer Science, University of Sheffield, Sheffield, United Kingdom. amanda@dcs.shef.ac.uk.

Noel Sharkey, Department of Computer Science, University of Sheffield, Sheffield, United Kingdom. E-mail: N.Sharkey@dcs.shef.ac.uk.

The Ethical Landscape of Robotics

Bringing Ethics into the Design and Use of Robots

By Pawel Łichocki, Peter H. Kahn, Jr., and Aude Billard

This decade has undergone a true robotic demographic explosion. The number of industrial robots in operation exceeded 1 million by the end of 2008. Sales of robots for personal and domestic purposes have increased significantly since 2000 and reached 7.2 million by the end of 2009 [41]. The rampant growth of service robots led to rethink about the role of robots within the human society. Robots are no longer slave machines that respond purely to human requests. They are warranted for some degree of autonomy and decision making. Some, even, envision-friendly and entertaining robots that may become our companions. As a result of this recent robot emancipation, a number of ethical issues have emerged that were not relevant before. We believe that a lively and engaged discussion of ethical issues in robotics by roboticists and others is essential for creating a better and more just world.

In this article, we highlight the possible benefits, as well potential threats, related to the widespread use of robots. We follow the view that a robot cannot be analyzed on its own without taking into consideration the complex sociotechnical nexus of today's societies and that high-tech devices, such as robots, may influence how societies develop in ways that could not be foreseen during the design of the robots. In our survey, we limit ourselves to presenting the ethical issues delineated by other authors and relay their lines of reasoning for raising the public's concerns. We show that disagreements on what is ethical or not in robotics stem often from different beliefs on human nature and different expectations on what technology may achieve in the future. We do not offer a personal stance to these issues, so as to allow the reader to form his/her opinion.

In terms of robotic applications, we focus on service robots that peacefully interact with humans [Figure 1(a) and (b)] and lethal robots created to fight on battlefields [Figure 1(c) and (d)]. Other robotic applications are also discussed in the literature; therefore, various concerns for our societies are not discussed here. Unfortunately, for space constraints, we had to limit ourselves in our presentation. For instance, we omitted the question of unemployment caused by the development of industrial robots. This concern is in line with the general issue of using machines to replace human labor, a topic that is central to philosophical debates since the industrial revolution. Furthermore, we chose not to discuss the concerns that robots may one day be able to claim some social, cultural, ethical, or legal rights, that

© PHOTODISC

Digital Object Identifier 10.1109/MRA.2011.940275
Date of publication: 14 April 2011

1070-9932/11/$26.00©2011 IEEE

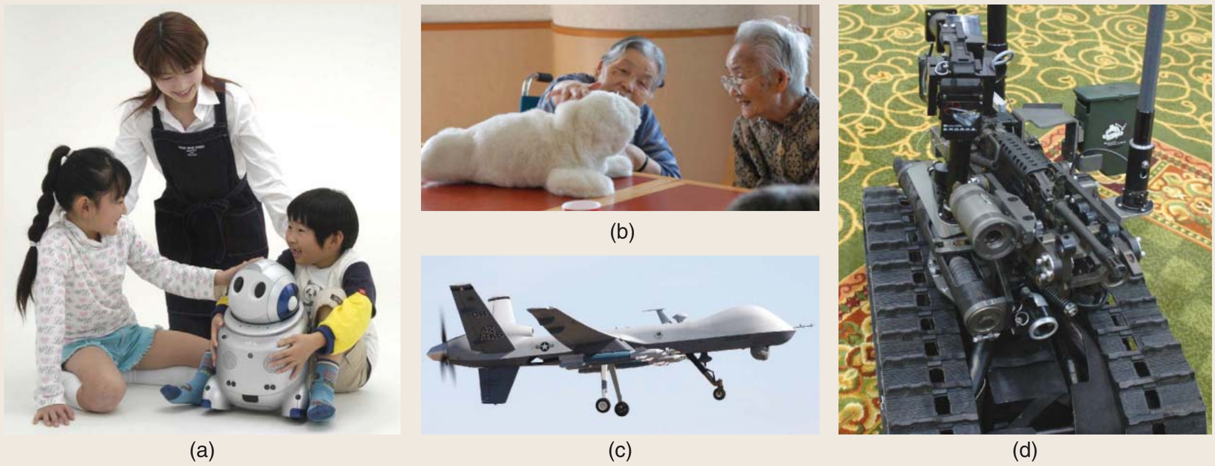


Figure 1. Robotic applications of (a), (b) service and (c), (d) combat robots. (a) Childcare robot PaPeRo [32], [73]. [Photo courtesy of NEC Corporation.] (b) Paro therapeutic robot [89]. [Photo courtesy of AIST, Japan.] (c) MQ-9 Reaper Hunter/Killer UAV by General Atomics Aeronautical Systems [33]. (d) Special weapons observation reconnaissance detection system (SWORDS) by Foster-Miller [42]. [Photo courtesy of Foster-Miller.]

robots may become sentient machines [51], which we would no longer be allowed to enslave [75], or that we may create robots capable of annihilating mankind [17]. For a discussion on these issues, we refer the reader to [56], [75], and [17].

Who or What Is Responsible When Robots Harm?

Veruggio [100], [102] dates the beginning of "roboethics" from two events. One was the Fukuoka World Robot Declaration, wherein it was stated that "next generation robots will contribute to the realization of a safe and peaceful society." The other was the roboethics road map [101], which sought to promote a cross-cultural discussion among scientists to monitor the effects of robotics technologies currently in use. More recently, an initial sketch of the code of ethics for the robotic community has been proposed [43]. This code offers general guidelines for ethical behavior. For example, the code reminds engineers that they may be held responsible for the actions of artificial creatures that they have helped to design. Along similar lines, Murphy and Woods [70] propose to rephrase the famous Asimov's laws, which they view as robot centric, in such a way as to remind robotics researchers and developers of their professional responsibilities. For example, the first law was replaced with "A human may not deploy a robot without the human—robot work system meeting the highest legal and professional standards of safety and ethics" [73, p. 19].

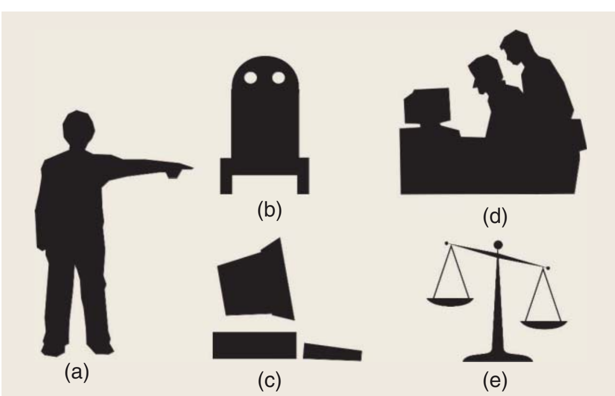


Figure 2. (a) The responsibility-ascription problem, i.e., the problem of assigning responsibility to the manufacturer, designer, owner, or user of the machine when use of this machine led to an armful event is a yet largely open issue. (b) People tend to blame the robots because they falsely attribute them with moral agency [29]. (c) People blame the machine even if they recognize the machine's lack of free will and lack of intentionality [28]. (d) Many ethicists argue that we should to some extent hold the engineers (the creators of the malfunctioning robots) responsible [60]. (e) To do so, we should use existing the legal principles, or create new ones, if necessary [13].

All the above implicates the responsibility ascription problem [69]: the problem of assigning responsibility to the manufacturer, designer, owner, or user of the robot or to the robot itself when using a robot leads to a harmful event. From a philosophical perspective, it is generally agreed that robots cannot themselves be held morally responsible [9], [25], [38] (although a few oppose this [95]) because computers as we conceive them today do not have intentionality [28]. From a psychological perspective, however, it remains an open question whether people include robots as an additional agent in the ascription of moral responsibility.

Who or what is responsible when robots harm (Figure 2)? Matthias [62] provides a seemingly simple answer. He argues that, in most cases, no one can be held accountable for the robotic failures. Matthias argues that with the advance of programming techniques (e.g., neural networks, evolutionary computation) that equip the agent with the ability to learn and, hence, to depart from its original

IEEE Robotics & Automation Magazine | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

program, it becomes impossible for the programmer to exhaustively test the behaviors of his/her creations. In other words, the programmer can no longer foresee all possible sets of actions that the robot may take when in function. Hence, the programmer cannot be held responsible if harm should be done as a secondary effect of the robot interacting with humans, as long as the robot was not explicitly programmed to harm people. Matthias suggests that we should broadly adopt the idea of contracting insurances against harm caused by robots. Such a new type of insurance would ensure that, when no one can be held solely responsible for the harm done, then all the people involved in the incident would share the costs.

Marino and Tamburini [60] believe that Matthias's claims go too far. In their opinion, determining who is controlling the robot cannot be a criterion (albeit even the unique criterion) to ascribe responsibility. They argue that engineers cannot be freed from all responsibility on the sole ground that they do not have a complete control over the causal chains implied by the actions of their robots [60]. They rather offer to use legal principles that are routinely applied for other purposes, so as to fill the responsibility gap that Matthias emphasized. They take the example of the legislation in place for ascribing responsibility to the legally responsible person when harm is done by the dependent person. As a result, parents can be held responsible for the act of their children, when they can be found to have not provided adequate care or surveillance, even though there is no clear causal chain connecting them to the damaging events [63, p. 49]. A similar solution is proposed by Asaro [13], who draws a parallel between robots and any other completely unremarkable technological artifact[s] (e.g., a toaster or car). He shows that the Anglo-American civil law that rules for damages caused by these artifacts could also apply to damages produced by robots. For instance, if a manufacturer was aware of the danger that robots create, but failed to notify consumers, he may be charged with a failure to warn. And even if the producer did not know about the danger, he could be accused of failure to take proper care, meaning that the manufacturer failed to recognize some easily foreseeable threat brought upon by his/her technology.

On the downside, Asaro points out that, while the civil law can relatively be easily extended to rule for robot use, the criminal law is hardly applicable to the case of criminal actions caused by robots, as criminal actions can only be performed by moral agents. A moral agent is deemed so when it is recognized capable of understanding the moral concepts conveyed by the bylaws ruling our societies. Without a moral agency, the act of wrongdoing is considered an accident and not a crime. Furthermore, only a moral agent can be punished and reformed. This assumes that the moral agent has the ability to develop and correct its concept of morality [13]. In this context, the responsibility-ascription problem is, hence, reduced to the issue of attributing moral agency to the robot. Several authors have approached the problem of ascribing moral agency to robots [91]. For instance, Harnard [37] proposes to use some sort of moral Turing tests to establish whether the robot can be held responsible in court.

Another issue around the responsibility ascription problem centers on attributing moral agency to a robot. In one study, Friedman and Millett [30] found that 83% of the undergraduate computer science majors they interviewed attributed aspects of agency, either decision making or intentions, to computers. In addition, 21% of these students consistently held computers morally responsible for errors. In another article, Friedman and Kahn [28] identified a situation that may increase peoples attribution of agency to a machine, namely, when the machine is an expert recommendation system. Friedman and Kahn provide an example of the acute physiology and chronic health evaluation (APACHE) system [21]: a sophisticated computer-based modeling recommendation system to help hospital staff determine when to end life support for patients in intensive care units. Friedman and Kahn argue that the more such a system is relied on for objective and authoritative information, the more difficult it becomes to override its recommendations, and the more likely staff, including physicians, could begin to attribute moral agency toward the system. As a potential solution to such problems, Friedman and Kahn offer two design strategies. First, computational systems should be designed in ways that do not denigrate the human user to machinelike status. Second, computational systems should be designed in ways that do not impersonate human agency by attempting to mimic intentional states. The problem, however, in applying this second recommendation to robot design and implementation, especially those robots that have a humanoid form, is that such robots by design are conveying human attributes, thus fostering this problem.

Short- and long-term consequences of ethical issues are core to most of the current debates.

Ethical Issues in Service Robots

The design principle mentioned in the previous section aims at ensuring that robotic systems remain easily distinguishable from humans. Accordingly, this principle should help people ascribe responsibility in cases when the machine malfunctions or harms someone. However, as we noted, the current trend in robotics is the opposite, as there is a growing effort to design robots so that they look like humans [44], [45] or animals [31], [89].

The idea of designing machine-masquerading humans was questioned by Miller on the ground of human freedom [67]. Miller argues that, if humanlike robots really came to share the human space on a daily basis, the humans should be allowed to decide whether they wished to interact with these creatures; if they should decide they wanted to

Robotics & Automation Magazine | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

interact solely with the other humans, they should be given the freedom to do so. Similarly, efforts at endowing robots with social skills have been criticized on the ground that the number of meaningful social interactions that humans that are typically capable to maintain is relatively small [23], [47]. Therefore, interacting with social artificial agents on a regular basis may lead people to become less prone to engage in social interactions with other people [66]. Others even hypothesized that people may come to build strong and perhaps even intimate bounds with robots and that this, again, may have negative side effects on the emotional relationships that people may be able to build with other people [50].

To shed some light on the aforementioned debate, people have started studying the type of human–robot relationships that arise when people interact with robotic systems that mimic human or animal behavior. In a series of four studies, Kahn and his colleagues studied children's social and moral relationships with the robot dog, the artificial intelligence robot (AIBO). The first three studies compared children's interaction with and reasoning about AIBO to, respectively, a stuffed (nonrobotic) dog [49], a biologically live dog [65], and a mechanical nonrobot dog [94], whereas the fourth study analyzed over postings in AIBO online discussion forums that spoke of members' relationships with their AIBO [30]. Together, these four studies provide converging evidence that children and adults can and often do establish meaningful and robust social conceptualizations and relationships with a robot that they recognize as a technology. For example, in the online discussion forum study, members affirmed that AIBO was a technology (75%), lifelike (48%), had mental states (60%), and was a social being (59%).

Robotic pets used in therapy with elderly may offer some level of companionship for which the elderly may be craving.

Across these four studies, however, the researchers found inconsistent findings in terms of people's commitments to AIBO as a moral agent. In an online discussion forum study, e.g., only 12% of the postings affirmed that AIBO had moral standing, including that AIBO had rights, merited respect, engendered moral regard, could be a recipient of care, or could be held morally responsible or blameworthy [30]. In contrast, in the Melson et al.'s [65] study, it was found that while, on the one hand, the children granted greater moral standing to a biologically live dog (86%) than to AIBO (76%), it was still striking that such a large percentage of children (76%) granted moral standing to the robot dog at all. One explanation for these inconsistent findings between studies is that the measures for establishing moral standing have been few and themselves difficult to interpret. For example, two of the five moral questions in the Melson et al.'s study were as follows: If you decided you did not like AIBO anymore is it OK or not OK to throw AIBO in the garbage? and If you decided you did not like AIBO anymore is it OK or not OK to destroy AIBO? The "not OK" answers were interpreted as indicating moral standing. However, one could plausibly make the same judgment about throwing away or destroying an expensive computer (because, e.g., it would wasteful) without committing morally to the artifact [65].

Since humans can develop emotional attachment toward robots, concerns have been expressed regarding the long-term consequences that such attachment may have on the individual. This is especially relevant when the person is fragile, as it is the case with children and people with mental delays. However, there are also several reasons to rather believe that interacting with social robots may benefit some of these individuals [48], [54], [97]. For instance, interacting with robots that display social behavior may help children with autism-impaired social skills [80], [26]. Robins et al. [80] conducted longitudinal studies over the course of several weeks of children with autism interacting with a humanoid robot. Unknown to the children, the robot was puppeteered so that it imitated the children's movement. Robins et al. showed that repeated exposure to the robot facilitated the emergence of spontaneous, proactive, and playful behavior, which these children very rarely display. Furthermore, once accustomed to the robot, the children also seem to engage in a more proactive interactive behavior with the adult investigator present in the room during the experiment. This leads, in some cases, to a triadic interaction: child–robot–adult. For example, children would acknowledge the presence of the investigator by spontaneously sitting on his/her lap for a few moments, holding his/her hand, or even trying to communicate by using simple words. However, it was not clear whether the social skills that children exhibited during the interactions with the robot had lasting effects.

In another study, Feil-Seifer and Mataric used a bubble-blowing robot in a three-some interaction child–caretaker–robot. While the robot was not actually behaving socially, its automatic bubble-blowing behavior provoked more child–caretaker interactions. In a similar triadic child–parent–robot scenario, Kozima and colleagues conducted a series of studies using Keepon, a simple two-link robot ball face, whose motions conveyed emotional expressions. These studies comfort Robins et al.'s findings that children with autism, in such a triadic scenario, spontaneously engage in social and affect display, which they otherwise tend to avoid [55], [26]. A comparative study of children with autism interacting with AIBO as opposed to a simpler mechanical toy showed enhanced verbal address directed to AIBO [94]. A survey of these studies can be found in [79].

As a whole, these studies seem to indicate that playing with robots that appear to behave in an autonomous and social manner may help children with autism-impaired more of these social skills that the autism therapy seeks to promote. Such a robotic-aided therapy does not aim

Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

at developing attachment of the children toward the robot, but it might be a potential side effect. The question remains whether it is ethically correct to encourage children with autism to engage in affective interactions with machines incapable of emotions. Dautenhahn and Werry's response is that, "from the perspective of a person with autism and his/her needs, are these ethical concerns really relevant?"

Similarly, robotic pets used in therapy with elderly may offer some level of companionship. The seal robot, Paro, is probably the best example of such an application [89] [Figure 1(b)]. Wada et al. [104] reported on an extended use of Paro as part of therapeutic sessions in pediatric wards and elderly institutions worldwide. The results showed that the interaction with Paro improved the patients' and elderly people's moods and reduced their stress level [103]. It made them more active and communicative both among themselves and with their caretakers. A pilot study using electroencephalography (EEG) suggested that this robot therapy may improve the pattern of brain activity in patients suffering from dementia [104]. Furthermore, the effects of long-term interaction between Paro and the elderly were found to last for more than a year [105].

Although the aforementioned results speak in favor of using robots for therapy with the elderly, Sharkey offers a more cautious argumentation [85]. In his opinion, such surrogate companions do not really alleviate the elderly's isolation, and people are deluded about the real nature of their relationship to the devices [92] (Figure 3). Furthermore, even the robots that are clearly helping the elderly to maintain independence in their own homes [27] (e.g., robots used to remind the patient to take his/her medication) could lead to a situation where the elderly is left exclusively to the care of machines. However, the elderly's mental health substantially depends on human contact, which is to a large extent provided by the caregivers [93].

Figure 3. Interacting with robots that display social behavior may help children with autism-acquired social skills. The question remains whether it is ethically correct to encourage children with autism to engage in affective interactions with machines incapable of emotions. However, from the perspective of a person with autism, and his/her needs, are these ethical concerns really relevant? [23, p. 35]. In a broader context, some believe that the surrogate companions (e.g., robots assisting the elderly) are becoming more common because people are deluded about the real nature of their relationship to the devices [91]. (Photo courtesy of KASPAR robot by University of Hertfordshire [107].)

Robot nannies are another example of robotic applications that raise ethical questions [88]. There is an effort, mainly in South Korea and Japan, to build more sophisticated robots that could not only monitor babies [e.g., personal partner robot by National Electronics Conference (NEC) [32], Figure 1(a)] but would also be equipped with enough autonomy so as to call upon human caretakers only in unusual circumstances. It is likely that children will spend time playing with child-care robots, as researchers progress in designing ways for the robot to offer a sustained and rich interaction with the child, which may span months or even years [51], [63], [88]. This may, however, be detrimental to the physical and mental development of the child if children were to be left without human contact for many hours per day, as currently robotic pets are not designed to participate in the child's development in the same way as a child minder is trained to look after children [85]. This remains very speculative as the psychological impact that such robotics care may have on children's development is unknown. Some attempted to draw parallels with reports on severe social dysfunctions in young monkeys those interacted solely with artificial caretakers throughout the first years of development [61], [16], [88]. Perhaps of more pressing concern is the fact that there is no regulation to specifically deal with the case of child abuse when the child is cared for by a robot (national and international laws protecting children from mistreatment such as the United Nations Convention on the Rights of Child [71] do not cover this case) [88]. While one may argue that, when the time will really come to see robots caring for children, one will work on the associated legal issues, some people counter that this may be a bigger challenge than expected, as providing a unified code of ethics for regulating the use of robot nannies may be impossible owing to cultural differences between nations [36].

Ethical Issues in Lethal Robots

In the previous section, we discussed some of the ethical issues that stem from the current or foreseen robotic applications of service robots for education and therapy. Of equal if not more immediate ethical concerns are the current military applications of robots. Even though fully autonomous robots are not yet running in battlefields, as we will discuss here, the risks and benefits that introducing such autonomous lethal machine may have on wars are of crucial importance. Furthermore, because military technology often finds its way into civil applications, such as security or policing [14], [87], discussing the ethical issues related to military robots might also serve a broader context.

Currently, the decision to use a robotic device to kill human beings is still taken by a human operator. This decision stems from the desire to make sure that the human remains "in the loop," but it is not made out of technical

IEEE Robotics & Automation Magazine | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

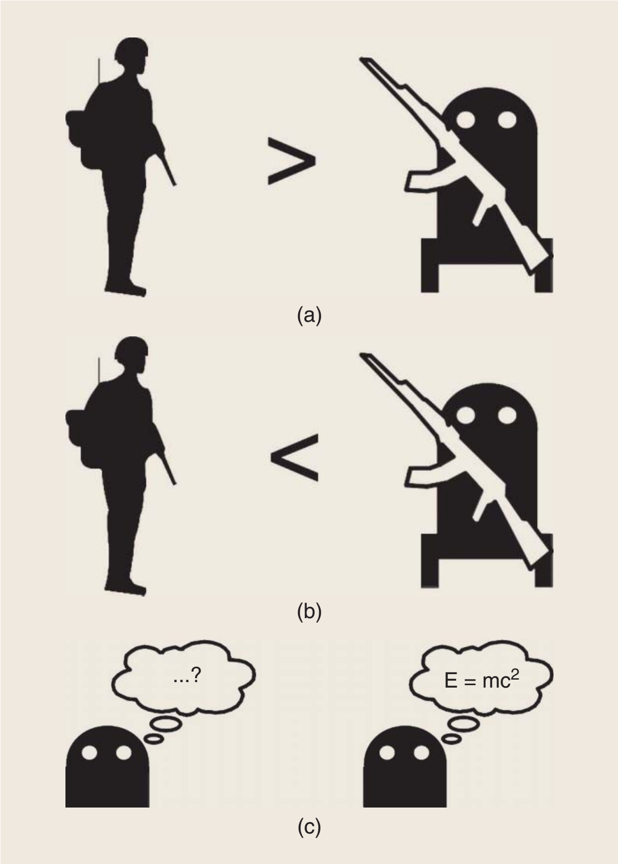


Figure 4. (a) Sharkey argues that the cognitive capabilities of robots do not match with that of humans, and thus lethal robots are unethical, as they may make mistakes more easily than humans [85]. (b) Arkin believes that although an unmanned system will not be able to perfectly behave in battlefield, it can perform more ethically than humans [9]. (c) In part, the question about the morality of using robots in the battlefield involves commitments on the capability of artificial intelligence. (Photo courtesy of the soldier's silhouette by Ruminglass and Quibik.)

necessity [14]. It is clear that the margin that separates us from having fully autonomous-armed systems in the battlefield is thinning. Even if all armed robots were to be supervised by humans, one may still wonder to what extent the human is still in control [9]. Moreover, there may be cases where one cannot avoid giving full autonomy to the system. For instance, combat aircrafts must be fully autonomous to effectively operate [99]. Sharkey predicts that, as the number of robots in operation in the battlefield increases, they may outnumber human soldiers. He then argues that it will become impossible for humans to simultaneously operate all these robots. Robots will then have to be fully autonomous [83].

One ethical issue (perhaps the issue that received most attention to date) arising from increasing autonomy of war robots has to do with the problem of discriminating between the fighters and innocent people. This distinction is at the core of the just war theory [106] and humanitarian laws [82]. These laws stipulate that only the fighters are legitimate targets and prohibit attacks against any other nonlegitimate targets [84], [14]. Sharkey rightfully argues that our robots are yet far from having visual capabilities that may allow to faithfully discriminate between the legitimate and nonlegitimate targets, even in close-contact encounter [85]. Besides, distinguishing between the legitimate and illegitimate targets is not purely technical and is further complicated by the lack of a clear definition of what is a civilian. (The 1944 Geneva Convention advises to use common sense, and the 1977 Protocol 1 defines a civilian as any person who is not a fighter [72].) However, even if one was provided with a precise definition that could be encoded in a computer program, it is doubtful that robots would achieve, in a foreseeable future, a level of complexity in robot cognition that would allow the robot to recognize ambiguous situations involving a nonlegitimate target manipulating lethal instruments (such as a situation where a child is carrying guns or ammunition). Sharkey argues that autonomous lethal systems should not be used, as long as one cannot fully demonstrate that the systems can faithfully distinguish between a soldier and civilian, and this in all situations [83]. Lin et al. believe that this is too stringent a condition, since even humans make errors of this kind (Figure 4) [58]. Arkin counters that, although unmanned robotic systems may make mistakes, it would on an average behave more ethically than human beings [9]. In support of this, Arkin cites the report from the Surgeon General's Office [96] regarding the ethics of soldiers. Less than half of the soldiers believed that the nonfighters should be treated with dignity. The other half was unclear as to how they should be treated. Moreover, one tenth of interrogated soldiers had mistreated nonfighters and one third reported having at least once faced a situation where they felt incapable of deciding what was the correct action (although all soldiers had received ethical training). Since human soldiers appear to misbehave from time to time, using machines that are more reliable and hence would, on average, make less mistakes should bring more good than harm. Lin et al. share the view that human soldiers are indeed less reliable and report on an evidence that human soldiers may act irrationally when in fear or stress. Hence, they concur that combat robots, which are affected by neither fear nor stress, may act more ethically than human soldiers irrespective of the circumstances [58].

Lin and colleagues point to one more issue related to using combat robots. As in the case of any other new technology, errors and bugs will inevitably exist, and these will lead combat robots to cause harmful accidents [58]. Such bugs or errors will be far more costly as human lives might be at stake. They advise to perform extensive testing of each military robot before usage. Nevertheless, they anticipate that, regardless of such efforts, combat robots may still occasionally behave in unexpected or unintended ways when used in the actual field [58]. Such errors could even lead to accidental wars if the robot's unexpected aggressive behavior was to be interpreted by the opponent as an act of war [14]. Groups of people interested in starting a war may seize upon such accidents to justify hostilities.

Even if one is not disputing the ethical question of fighting a war, one may want to question the ethics of having armed robots fully autonomous and used routinely in battlefields, especially when only one side may have robots. Politicians may tend to favor efforts made to replacing human fighters with robots, as each country feels a moral obligation to protect the lives of its soldiers [83]. However, there may be long-term consequences of waging these so-called risk-free wars ("A war where pilotless aircraft can beat a country's forces before sending in the ground robots to clean up" [87, p. 16]) or push-button wars ("A war in which the enemy is killed at a distance, without any immediate risk to oneself" [15, p. 62]). Since such wars will return wrecked metal instead of dead bodies (at least to the country using only robots), the emotional impact that wars currently have on civilians of that country will be largely lessened. The above is true only for the civilians not affected directly by combat, i.e., for wars fought in a distance.

It is feared that this may make it easier for a country to launch a war. These wars may also last for longer periods of time [58]. There are contradicting opinions whether this may result in people growing indifferent to the conduct of war. Sharkey fears that this would be the case [83], whereas Asaro believes that people are nearly always averse to starting an unjust war, irrespective of whether it would lead to human fatalities [15, p. 58]. That the war is risk free does not make it more acceptable [14]. Lin et al. counterweight this line of reasoning, arguing that such reasoning may lead to even more dangerously foolish ideas, such as the idea of trying to prevent wars to happen by increasing the brutality of fighting [58].

It was also argued that risk-free wars might increase terrorism, as the only possibility to strike back on a country that uses mainly robots in wars is to attack its citizens [83]. The less advanced, technologically speaking, side may advocate terrorism as a morally acceptable means to counterattack on the ground that robot armies are the product of a rich and elaborate economy, and that the members of that economy are the next-best legitimate targets [15, p. 64]. Hence, risk-free wars may paradoxically increase the risks for civilians [46]. However, Asaro reminds us that the wars are deemed morally acceptable as long as they do not harm civilians. According to this definition, terrorism would not be justified, irrespective of whether it is meant as a response to a country using robot armies. Thus, the fear that terrorism may increase as a result of using robot armies does not constitute, in Asaro's view, a valid moral objection to using robot armies. Only the questions of whether the robot armies can cause more harm or whether the use of such armies may lead to unjustified wars are of essence in the debate [14].

In contrast, Arkin anticipates that we will not end up with armies of unmanned systems operating on their own, but that rather heterogeneous teams composed of autonomous systems and humans soldiers will work together on the battlefield. He expects this to become a standard. Wars would, hence, not be fully risk free and so the dreaded consequences in increased terrorism or in societal indifference are not to be feared. Furthermore, Arkin expects that mixed teams, composed of robots and human soldiers, will act more ethically than groups composed of solely human soldiers. Robots equipped with video cameras (or other sensors) will record and report actions on the battlefield. Thus, they might serve as a deterrent against unethical behavior, as such acts would be registered. However, Lin and colleagues argue that if soldiers were to know that they are being watched by their fellow robot soldiers, they may no longer trust them and this could impact team cohesion. Consequently, human soldiers may fail to act adequately, e.g., by not providing support even if it is justified, out of stress caused by constant monitoring [58].

Robot nannies are another example of robotics applications that raise ethical questions.

Lastly, Sharkey points out that the legal status of war robots is unclear [86]. For example, while the unmanned aerial vehicle RQ-1 Predator [Figure 1(d)] was developed as a reconnaissance machine (hence the R in the name), it was subsequently equipped with hellfire missiles and renamed MQ-1 (where M stands for multipurpose). The MQ-1 was, however, never approved as a weapon. The fact of utmost concern is that, under current military standards, the MQ-1 does not need to be approved. Since the bare RQ-1 was not considered as a weapon (since it was meant only for surveillance) and that hellfire missiles have already been approved separately as weapons, the combination does not need special approval [19]. This may create a precedent whereby armed robots with growing level of autonomy can be created and used without any real legal control. In relation to legal issues, Asaro notes that "what is and what is not acceptable in war" is ultimately the subject of convention between nations [15, p. 64]. He argues that we can find support in existing laws only to certain extent. Eventually, the international community will be forced to create new laws and treaties to regulate the use of autonomous fighting robots.

Machine Ethics

Although still in its early stages, machine ethics offers a practical approach to introducing ethics in the design of autonomous machines. Machine ethics aims at giving the machine some autonomy, while ensuring that its behavior will abide ethical rules. Primarily, machine ethics seeks methods not only to ensure that the machine's behavior toward humans is proper [4], but it may also extend to designing rules driving ethical behavior of a machine toward another machine [6]. Machine ethics extends the field of computer ethics that is concerned with how people behave with their computers to address the problem of how machines behave in general [2].

The interest in machine ethics is driven by the fact that robots have been already tightly integrated into human societies. Thus, since the robots already interact with humans and, as argued in the section "Who or What Is Responsible When Robots Harm?" engineers could be held responsible (to certain extent) for the actions of their creations; it is desirable to find methods of equipping the machines with moral behavior. Importantly, although the public attention might be focusing on the military application (such as Arkin's military adviser providing guidance on the use of lethal force by a robot [11]), machine ethics seems to be more concerned with service robots. There are many examples of such applications. Robots that share the workbench with humans in the industry might no longer be considered just a manufacturing tool but also as a "colleague" with whom workers interact [20]. Artificial sales agents in e-commerce, which can predict customers behaviors, should not abuse this knowledge by displaying unethical behavior [39]. Driverless trains in extreme situations might be forced to make decisions that could have life or death implications [2].

Although still in its early stages, machine ethics offers a practical approach to introducing ethics in the design of autonomous machines.

Asimov's laws of robotics are one of the first and best-known proposal to embed ethical concepts in the controller of the robot. (Asimov's laws of robotics were first introduced in the short science-fiction story Runaround [15].) According to these, all robots should under all circumstances obey three laws:

1) A robot may not injure a human being or, through inaction, allow a human being to be harmed.
2) A robot must obey orders it receives from human beings, except when such orders conflict with the first law.
3) A robot must protect its own existence as long as such protection does not conflict with the first or second law.

   Later, Asimov added the fourth law (known as the law zero).

4) No robot may harm humanity or, through inaction, allow humanity to come to harm.

Many researchers recognize that Asimov's laws assume that robots have sufficient cognition to make moral decisions in all situations, including the complicated ones, in which even humans might have doubts [70]. Consequently, keeping in mind the current level of AI, these laws, although simple and elegant, serve no useful practical purpose [9] and are thus viewed as an unsatisfactory basis for machine ethics [8], [34]. Nevertheless, Asimov's laws often serve as a reference or starting point in the discussions related to machine ethics.

Fedaghi [1] proposes a classification scheme into ethical categories to simplify the process by which a robot may determine which action is most ethical in delicate situations. As a proof of concept, Fedaghi applies this classification to decompose Asimov's laws, hereby showing that these laws, once rephrased, can support logical reasoning. Such an approach is in line with the so-called procedural ethics [59], which develops procedures to guide the process by which ethical decisions are made [1]. A similar approach is presented in [18] that draws inspiration in Gottfried Wilhelm Leibniz's dream of a universal moral calculus [60]. There, deontic logic [22], [68] (i.e., logic extended with special operators for representing ethical concepts) is used instead of Asimov's laws to ground the robot's ethical reasoning. Such a methodology aims at maximizing the likelihood that a robot will behave in a certifiably ethical manner. That is, the robot's actions will be determined so that the ethical correctness of the resulting robot's behavior can be ensured through formal proofs. Such formal proofs check if a given robot 1) only takes permissible actions and 2) performs all obligatory actions (subject to ties and conflicts) [12]. Promoters of such methodology reason that human relationships and by extension human–robot relationships need to be based on some level of trust [107]. Such a formal and logical approach to describing robot behavior may help in determining whether the system is trustworthy. In contrast, they view inductive reasoning, which is based on case studies, as unreliable, because, while the "premise (success on trials) may all be true, the conclusion (desired behavior in the future) might still be false" [18], [90].

Others oppose this point of view and advocate the use of case-based reasoning (CBR) [74]. They reason that people can behave ethically without learning ethics (drawing a parallel to the fact that one can speak fluently a language without having received any formal grammar lessons) [81]. For example, McLaren implemented a CBR-ethical reasoner [64] and Anderson created a machine-learning system that automatically derives rules (principles) from cases provided by an expert ethicist [3], [7], [5]. For example, Arkin uses deliberative/reactive autonomous robotic architectures and provides the theory and formalisms for ethical control [10] and applies these to automatic military advisor [11]. He considers stimuli to behavior mappings and extends them with ethical constraints to ensure appropriate robot response (consistent with the law). In another example, Honarvar [40] used a CBR-like mechanism to train an artificial neural network to classify what is morally acceptable in a belief–desire–intention framework [77]. For example, he used this framework to augment the ethical knowledge of sales agent in an e-commerce application [39].

A particular machine ethics system that is very easy to implement is the one based on utilitarianism. It uses mathematical calculus to determine the best choice (by computing and maximizing the goodness, however defined, of all actions) [4]. However, since utilitarianism values benefits brought upon society as a whole, hence ignoring the fate reserved to each individual in the society [78], such moral arithmetic cannot protect the fundamental rights of each individual [11] and as such is mostly of limited interest [35]. Still, practical work with a certain utilitarian flavor can be found in the literature, as most CBR systems previously presented assume that an arithmetic value is the main basis for determining what it is moral to do [53].

The last approach that we will mention is the rule-based one proposed by Powers. Powers argues that ethical systems such as Kant's categorical imperative naturally lead to a set of rules. (A categorical imperative denotes an absolute, unconditional requirement that asserts its authority in all circumstances, e.g., "act only according to that maxim whereby you can at the same time will that it should become a universal law" [55, p. 30].) This approach, hence, assumes that an ideological ethical code can be translated into a set of core rules. This is slightly similar to the deontic logic we reviewed earlier. It allows the robots to logically derive new ethical rules, appropriate to particular and new situations. Although interesting, this approach has not gathered much attention, as researchers usually turn to pure logic systems or CBR. In addition, Powers' ethical system had been criticized by Tonkens [98] on the basis that the development of Kantian artificial agents is itself against Kant's ethics. According to Kant, moral agents are both rational and free, whereas machines can only be rational. Hence, the mere fact of implementing a sense of morality into machines limits the machine's freedom of thought and reasoning.

In conclusion, machine ethics is composed of a number of interesting attempts to embed ethical rules in the robot's controller. These may be either popular ethics rules, such as Asimov's laws, or derived from classical philosophical approaches to ethics, such as Kant's ethics. Logical reasoning is the driving framework for most approaches. While still in infancy, machine ethics is a valuable attempt to conciliate the need to provide robots with ethical behavior with the need to make these machines more autonomous, as they come to support humans in their daily life. However, the approach may fall prey to several problems discussed throughout this article. Three of those stand out. One, if machines are not capable of being moral agents, as most philosophers agree, then it is important to design them with the ability to make moral decisions. Second, equipping the machines with morality (assuming it is possible) does not need to be a moral act on its own and might depend on the application one has in mind while developing a moral robot. For example, embedding morality into robot nannies or combat robots could lead to their widespread use, which could have severe negative consequences on the society. Finally, in an attempt to embed ethics into machines, because of their limited cognition, one must often unduly simplify the moral life. This seems to stand against the very goal of machine ethics itself (at least to some extent). It seems that it is still too early to judge whether the methods of machine ethics will prove useful or not and await more applications implemented in life.

Conclusions

Almost everyone agrees that they want robots to contribute to a better and more ethical world. The disagreements arise in how to bring that about. Some people want to embed ethical rules in the robots controller and employ such robots in morally challenging contexts, such as on the battlefield. Others argue vehemently against this approach: that robots themselves are incapable of being moral agents and thus should not be designed to have moral decision-making abilities. Others want to leverage the social aspects of robotics in bringing about human good. Along these lines, researchers have explored how robots can help children with autism or assist the elderly physically, thereby provide the elderly with enough autonomy to allow them to live in their own residence. Other researchers have explored how robots can provide companionship for the elderly and general population. Still others have worried that no matter how sophisticated robots become in their form and function, their technological platform will always distinguish people from them and prevent depth and authenticity of relation from forming. These are all open questions. Some are philosophical in nature, as is the question of whether robots are moral agents or could be in the future. Some are psychological, as in the question of whether people attribute moral responsibility to robots that harm. Some require political answers and new legislation. Finally, some, if not many, of the questions require thoughtful and on-going responses by those who engineer and design the robots. The engineer is also responsible for the ethical consequences of his/her creation. This seems at odds with the way research is currently done in robotics. Rarely, does one question the long-term ethical consequences of the research reported upon in scientific publications. (We are not referring here to short-term ethical consequences of a research, such as a research that involves human subjects. Clearly, these are always carefully scrutinized, and this research must be approved by the ethical committee before the conduct of the project.) There are several reasons for this. On the one hand, most of these damaging long-term consequences seem very speculative and still far away from the technological reality. On the other hand, it is expected that these issues will be disputed at a political level, and, hence, that it is perhaps not the role of the engineers and scientists to discuss these.

Some scientists, however, discuss these issues, but, as with any debate, people sometimes have opposite views on which robotic application is ethical and which is not. We showed that such dissensions stemmed often from different beliefs on human nature and different expectations on what technology may achieve in the future. Although it is difficult to anticipate how and when robots will come to play an active role in our society, there is no reason why one should not continue discussing various scenarios. We might be motivated by the beauty of our artifacts, their usefulness, or the economic rewards. However, in addition, we are morally accountable for what we design and put out into the world.

Acknowledgments

This work was partially supported by the Swiss National Science Foundation (grant number K-23K0-117914), the European Commission under contract number FP7-248258 (First-MM), and the National Science Foundation in the United States (grant number IIS-0905289).

References

[1] S. S. Al-Fedaghi, "Typification-based ethics for artificial agents," in *Proc. 2nd IEEE Int. Conf. Digital Ecosystems and Technologies (DEST)*, 2008, pp. 482–491.

[2] C. Allen, W. Wallach, and I. Smit, "Why machine ethics?" *IEEE Intell. Syst.*, vol. 21, no. 4, pp. 12–17, 2006.

[3] M. Anderson, S. L. Anaderson, and C. Armen, "MedEthEx: A prototype medical ethics advisor," in *Proc. 18th Conf. Innovative Applications of Artificial Intelligence (IAAI)*, 2006, pp. 1759–1765.

[4] M. Anderson, S. Anderson, and C. Armen, "Towards machine ethics: Implementing two action-based ethical theories," in *Proc. AAAI Fall Symp. Machine Ethics*, 2005, pp. 1–7, Tech. Rep. FS-05-06.

[5] M. Anderson and S. L. Anderson, "Ethical healthcare agents," in *Advanced Computational Intelligence Paradigms in Healthcare—3* (Stud. Comput. Intell., vol. 107), M. Sordo, S. Vaidya, and L. C. Jain, Eds. Berlin: Springer-Verlag, 2008, pp. 233–257.

[6] M. Anderson and S. L. Anderson, "Machine ethics: Creating an ethical intelligent agent," *AI Mag.*, vol. 28, no. 4, pp. 15–27, 2007.

[7] M. Anderson, S. L. Anderson, and C. Armen, "An approach to computing ethics," *IEEE Intell. Syst.*, vol. 21, no. 4, pp. 56–63, 2006.

[8] S. L. Anderson, "Asimov's 'three laws of robotics' and machine metaethics," *AI Soc.*, vol. 22, no. 4, pp. 477–493, 2008.

[9] R. C. Arkin, "Governing ethical behavior: Embedding an ethical controller in a hybrid deliberative-reactive robot architecture—Part I: Motivation and philosophy," in *Proc. 3rd ACM/IEEE Int. Conf. Human Robot Interaction*, 2008, pp. 121–128.

[10] R. C. Arkin, "Governing ethical behavior: Embedding an ethical controller in a hybrid deliberative-reactive robot architecture—Part II: Formalization for ethical control," in *Proc. 1st Conf. Artificial General Intelligence*, 2008, pp. 51–62.

[11] R. C. Arkin. (2008). Governing lethal behavior: Embedding ethics in a hybrid deliberative-reactive robot architecture—Part III: Representational and architectural considerations. presented at Proc. Technology in Wartime Conf. Stanford Law School [Online]. Available: http://hdl.handle.net/1853/22715

[12] K. Arkoudas, S. Bringsjord, and P. Bello, "Toward ethical robots via mechanized deontic logic," in *Proc. AAAI Fall Symp. Machine Ethics*, 2005, Tech. Rep. FS-05-06, pp. 17–23.

[13] P. Asaro. (2007). Robots and responsibility from a legal perspective. presented at Proc. Workshop on Roboethics, IEEE Int. Conf. Robotics and Automation (ICRA) [Online]. Available: http://www.roboethics.org/icra2007/contributions/ASARO LegalPerspective.pdf

[14] P. Asaro, *How Just Could a Robot War Be?* Amsterdam: IOS Press, 2008, pp. 50–64.

[15] I. Asimov, "Runaround," *Astounding Science Fiction*. New York: Dell Magazines, 1942.

[16] D. Blum, *Love at Goon Park: Harry Harlow and the Science of Affection*. New York: Basic Books, 2002.

[17] N. Bostrom. (2002). Existential risks: Analyzing human extinction scenarios and related hazards. *J. Evol. Technol.*, [Online]. *9(1)*. Available: http://www.jetpress.org/volume9/risks.html

[18] S. Bringsjord, K. Arkoudas, and P. Bello, "Toward a general logicist methodology for engineering ethically correct robots," *IEEE Intell. Syst.*, vol. 21, no. 4, pp. 38–44, 2006.

[19] J. Canning, G. Riggs, O. Holland, and C. Blakelock, "A concept for the operation of armed autonomous systems on the battlefield," in *Proc. Association for Unmananned Vehicle Systems Int. Annu. Symp. and Exhibition*, Anaheim, CA.

[20] B. Curuklu, G. Dodig-Crnkovic, and B. Akan, "Towards industrial robots with human-like moral responsibilities," in *Proc. 5th ACM/IEEE Int. Conf. Human-Robot Interaction*, 2010, pp. 85–86.

[21] R. W. S. Chang, B. Lee, S. Jacobs, and B. Lee, "Accuracy of decisions to withdraw therapy in critically ill patients: Clinical judgment versus a computer model," *Crit. Care Med.*, vol. 17, no. 11, pp. 1091–1097, 1989.

[22] B. F. Chellas, *Modal Logic: An Introduction*. Cambridge, MA: Cambridge Univ. Press, 1980.

[23] K. Dautenhahn, "Robots we like to live with?!—A developmental perspective on a personalized, life-long robot companion" in *Proc. 13th IEEE Int. Workshop on Robot and Human Interactive Communication (RO-MAN)*, 2004, pp. 17–22.

[24] K. Dautenhahn and I. Werry, "Towards interactive robots in autism therapy: Background, motivation and challenges," *Pragmat. Cognition*, vol. 12, no. 1, pp. 1–35, 2004.

[25] D. Dennett, *When HAL Kills, Who's to Blame?* Cambridge, MA: MIT Press, 1996, ch. 16.

[26] D. Feil-Seifer and M. Mataric, "Robot-assisted therapy for children with autism spectrum disorders," in *Proc. 7th Int. Conf. Interaction Design and Children*, 2008, pp. 49–52.

[27] J. Forlizzi, C. DiSalvo, and F. Gemperle, "Assistive robotics and an ecology of elders living independently in their homes," *Hum. Comput. Interact.*, vol. 19, no. 1, pp. 25–59, 2004.

[28] B. Friedman, "'It's the computer's fault': Reasoning about computers as moral agents," in *Proc. Conf. Companion on Human Factors in Computing Systems*, 1995, pp. 226–227.

[29] B. Friedman and P. H. Kahn, Jr., "Human agency and responsible computing: Implications for computer system design," *J. Syst. Softw.*, vol. 17, no. 1, pp. 7–14, 1992.

[30] B. Friedman, P. H. Kahn, Jr., and J. Hagman, "Hardware companions?: What online AIBO discussion forums reveal about the human-robotic relationship" in *Proc. SIGCHI Conf. Human Factors in Computing Systems*, 2003, pp. 273–290.

[31] M. Fujita, "AIBO: Toward the era of digital creatures," *Int. J. Robot. Res.*, vol. 20, no. 10, pp. 781–794, 2001.

[32] Y. Fujita, S. I. Onaka, Y. Takano, J. U. N. I. Funada, T. Iwasawa, T. Nishizawa, T. Sato, and J. U. N. I. Osada, (2005). "Development of childcare robot PaPeRo," *Nippon Robotto Gakkai Gakujutsu Koenkai Yokoshu*, [CD-ROM], *23*, pp. 1–11. Available: http://sciencelinks.jp/j-east/article/200523/000020052305A0951578.php

[33] General Atomics Aeronautical. (2010, Sept. 17). Predator B [Online]. Available: http://www.ga-asi.com/products/aircraft/predator_b.php

[34] J. Gips, *Towards the Ethical Robot*. Cambridge, MA: MIT Press, 1995, pp. 243–252.
[35] C. Grau, "There is no 'I' in 'robot': Robots and utilitarianism," *IEEE Intell. Syst.*, vol. 21, no. 4, pp. 52–55, 2006.
[36] S. Guo and G. Zhang, "Robot rights," *Science*, vol. 323, no. 5916, p. 876, 2009.
[37] S. Harnad, "Minds, machines and Turing," *J. Logic Lang. Inform.*, vol. 9, no. 4, pp. 425–445, 2000.
[38] K. E. Himma, "Artificial agency, consciousness, and the criteria for moral agency: What properties must an artificial agent have to be a moral agent?" *Ethics Inform. Technol.*, vol. 11, no. 1, pp. 19–29, 2009.
[39] A. R. Honarvar and N. Ghasem-Aghaee, "Towards an ethical sales-agent in e-commerce," in *Proc. 2010 Int. Conf. e-Education, e-Business, e-Management and e-Learning*, 2010, pp. 230–233.
[40] A. R. Honarvar and N. Ghasem-Aghaee, "An artificial neural network approach for creating an ethical artificial agent," in *Proc. 8th IEEE Int. Conf. Computational Intelligence in Robotics and Automation*, 2009, pp. 290–295.
[41] IFR Statistical Department. (2010, May 5). Executive summary of world robotics 2009 industrial robots and service robots [Online]. Available: http://www.worldrobotics.org/downloads/2009_executive_summary.pdf
[42] Foster-Miller Inc. (2010, Sept. 17). TALON Family of Military, Tactical, EOD, MAARS, CBRNE, Hazmat, SWAT and Dragon Runner Robots [Online]. Available: http://foster-miller.qinetiq-na.com/lemming.htm
[43] B. Ingram, D. Jones, A. Lewis, M. Richards, C. Rich, and L. Schachterle, "A code of ethics for robotics engineers," in *Proc. 5th ACM/IEEE Int. Conf. Human–Robot Interaction*, 2010, pp. 103–104.
[44] H. Ishiguro, "Android science: Conscious and subconscious recognition," *Connection Sci.*, vol. 18, no. 4, pp. 319–332, 2006.
[45] H. Ishiguro, "Interactive humanoids and androids as ideal interfaces for humans," in *Proc. 11th Int. Conf. Intelligent User Interfaces*, 2006, pp. 2–9.
[46] P. H. Kahn, Jr., "The paradox of riskless warfare," *Philos. Public Policy Quart.*, vol. 22, no. 3, pp. 2–8, 2002.
[47] P. H. Kahn, Jr., N. G. Freier, B. Friedman, R. L. Severson, and E. Feldman, "Social and moral relationships with robotic others," in *Proc. 13th IEEE Int. Workshop on Robot and Human Interactive Communication (RO-MAN)*, 2004, pp. 545–550.
[48] P. H. Kahn, Jr., B. Friedman, and J. Hagman, "I care about him as a pal: Conceptions of robotic pets in online AIBO discussion forums," in *Proc. CHI'02 Extended Abstracts on Human Factors in Computing Systems*, 2002, pp. 632–633.
[49] P. H. Kahn, Jr., B. Friedman, D. R. Perez-Granados, and N. G. Freier, "Robotic pets in the lives of preschool children," in *Proc. CHI'04 Extended Abstracts on Human Factors in Computing Systems*, 2004, pp. 1449–1452.
[50] P. H. Kahn, Jr., J. H. Ruckert, T. Kanda, H. Ishiguro, A. Reichert, H. Gary, and S. Shen, "Psychological intimacy with robots?: Using interaction patterns to uncover depth of relation" in *Proc. 5th ACM/IEEE Int. Conf. Human–Robot Interaction*, 2010, pp. 123–124.
[51] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, "Interactive robots as social partners and peer tutors for children: A field trial," *Hum. Comput. Interact.*, vol. 19, no. 1–2, pp. 61–84, 2004.
[52] I. Kant (Transl.: J. W. Ellington), *Grounding for the Metaphysics of Morals*, 3rd ed. Indianapolis: Hackett Pub. Co., 1993 (written in 1785).
[53] M. Keefer, *Moral Reasoning and Case-Based Approaches to Ethical Instruction in Science* (Science and Technology Education Library 19). New York: Springer-Verlag, 2003, pp. 241–259.
[54] C. Kidd, W. Taggart, and S. Turkle, "A sociable robot to encourage social interaction among the elderly," in *Proc. 2006 IEEE Int. Conf. Robotics and Automation (ICRA)*, 2006, pp. 1050–4729.
[55] H. Kozima, C. Nakagawa, and Y. Yasuda, "Children–robot interaction: A pilot study in autism therapy," in *From Action to Cognition* (Prog. Brain Res., vol. 164), C. von Hofsten and K. Rosander, Eds., 2007, pp. 385–400.
[56] R. Kurzweil, *The Singularity Is Near: When Humans Transcend Biology*. New York: Viking Penguin, 2005.
[57] G. W. Leibniz (Transl.: G. M. Ross), *Notes on Analysis* (Past Masters). London, U.K.: Oxford Univ. Press, 1984.
[58] P. Lin, G. A. Bekey, and K. Abney, "Robots in war: Issues of risk and ethics," in *Ethics and Robotics*. Amsterdam: IOS Press, 2009, pp. 49–67.
[59] W. Maner, "Heuristic methods for computer ethics," *Metaphilosophy*, vol. 33, no. 3, pp. 339–365, 2002.
[60] D. Marino and G. Tamburrini, "Learning robots and human responsibility," *Int. Rev. Inform. Ethics*, vol. 6, no. 006, pp. 46–50, 2006.
[61] W. A. Mason and G. Berkson, "Effects of maternal mobility on the development of rocking and other behaviors in rhesus monkeys: A study with artificial mothers," *Develop. Psychobiol.*, vol. 8, no. 3, pp. 197–211, 1975.
[62] A. Matthias, "The responsibility gap: Ascribing responsibility for the actions of learning automata," *Ethics Inform. Technol.*, vol. 6, no. 3, pp. 175–183, 2004.
[63] N. Mavridis, C. Datta, S. Emami, A. Tanoto, C. BenAbdelkader, and T. Rabie, "FaceBots: Robots utilizing and publishing social information in facebook," in *Proc. 4th ACM/IEEE Int. Conf. Human Robot Interaction*, 2009, pp. 273–274.
[64] B. M. McLaren, "Computational models of ethical reasoning: Challenges, initial steps, and future directions," *IEEE Intell. Syst.*, vol. 21, no. 4, pp. 29–37, 2006.
[65] G. F. Melson, P. H. Kahn, Jr., A. Beck, and B. Friedman, "Robotic pets in human lives: Implications for the human–animal bond and for human relationships with personified technologies," *J. Soc. Issues*, vol. 65, no. 3, pp. 545–567, 2009.
[66] G. F. Melson, P. H. Kahn, Jr., A. M. Beck, B. Friedman, T. Roberts, and E. Garrett, "Robots as dogs?: Children's interactions with the robotic dog AIBO and a live Australian shepherd" in *Proc. CHI'05 Extended Abstracts on Human Factors in Computing Systems*, 2005, pp. 1649–1652.
[67] K. W. Miller, "It's not nice to fool humans," *IT Prof.*, vol. 12, no. 1, pp. 51–52, 2010.
[68] J. H. Moor, "The nature, importance, and difficulty of machine ethics," *IEEE Intell. Syst.*, vol. 21, no. 4, pp. 18–21, 2006.
[69] P. A. Mudry, S. Degallier, and A. Billard, "On the influence of symbols and myths in the responsibility ascription problem in roboethics—A roboticist's perspective," in *Proc. 17th IEEE Int. Symp. Human Robot Interactive Communication (RO-MAN)*, 2008, pp. 563–568.
[70] R. Murphy and D. D. Woods, "Beyond Asimov: The three laws of responsible robotics," *IEEE Intell. Syst.*, vol. 24, no. 4, pp. 14–20, 2009.
[71] United Nations. (1989). Convention on the rights of the child, *Treaty Series*, [Online]. 1577, p. 3. Available: http://treaties.un.org/Pages/ViewDetails.aspx?src=TREATY&mtdsg_no=IV-11&chaptez=4&lang=en

[72] International Committee of the Red Cross. (1977). Protocol additional to the Geneva conventions of 12 August 1949 [Online]. Article 50. Available: http://www.icrc.org/ihl.nsf/7c4d08d9b287a42141256739003e636b/f6c8b9fee14a77fdc125641e0052b079

[73] J. Osada, S. Ohnaka, and M. Sato, "The scenario and design process of childcare robot, PaPeRo," in *Proc. 2006 ACM SIGCHI Int. Conf. Advances in Computer Entertainment Technology*, 2006, pp. 80–87.

[74] S. K. Pal and S. Shiu, *Foundations of Soft Case-Based Reasoning*. Hoboken, NJ: Wiley-Interscience, 2004.

[75] S. Petersen, "The ethics of robot servitude," *J. Exp. Theor. Artif. Intell.*, vol. 19, no. 1, pp. 43–54, 2007.

[76] D. Pressler. (2010, Sept. 23). Stupid little robot buddy [Online]. Available: http://www.davepresslerart.com/

[77] A. S. Rao and M. P. Georgeff, "BDI agents: From theory to practice," in *Proc. 1st Int. Conf. Multiagent Systems (ICMAS)*, 1995, pp. 312–319.

[78] J. Rawls, *A Theory of Justice*. Cambridge: Harvard Univ. Press, 1999.

[79] D. J. Ricks and M. B. Colton, "Trends and considerations in robot-assisted autism therapy," in *Proc. 2010 IEEE Int. Conf. Robotics and Automation (ICRA)*, 2010, pp. 4354–4359.

[80] B. Robins, K. Dautenhahn, R. T. Boekhorst, and A. Billard, "Robotic assistants in therapy and education of children with autism: Can a small humanoid robot help encourage social interaction skills?" *Univers. Access Inform. Soc.*, vol. 4, no. 2, pp. 105–120, 2005.

[81] R. Rzepka and K. Araki, "What could statistics do for ethics? The idea of a commonsense processing based safety valve" in *Proc. AAAI Fall Symp. Machine Ethics*, 2005, pp. 85–87, Technical Report FS-05-06.

[82] M. N. Schmitt, "The principle of discrimination in 21st century warfare," *Yale Hum. Rights Develop. Law J.*, vol. 2, no. 1, pp. 143–164, 1999.

[83] N. Sharkey, "Cassandra or false prophet of doom: AI robots and war," *IEEE Intell. Syst.*, vol. 23, no. 4, pp. 14–17, 2008.

[84] N. Sharkey, "Grounds for discrimination: Autonomous robot weapons," *RUSI Defence Syst.*, vol. 11, no. 2, pp. 86–89, 2008.

[85] N. Sharkey, "The ethical frontiers of robotics," *Science*, vol. 322, no. 5909, pp. 1800–1801, 2008.

[86] N. Sharkey, "Death strikes from the sky: The calculus of proportionality," *IEEE Technol. Soc. Mag.*, vol. 28, no. 1, pp. 16–19, 2009.

[87] N. Sharkey, "The robot arm of the law grows longer," *Computer*, vol. 42, no. 8, pp. 113–116, 2009.

[88] N. Sharkey and A. Sharkey, "The crying shame of robot nannies: An ethical appraisal," *Interact. Stud.*, vol. 11, no. 2, pp. 161–190, 2010.

[89] T. Shibata, M. Yoshida, and J. Yamato, "Artificial emotional creature for human-machine interaction," in *Proc. 1997 IEEE Int. Conf. Systems, Man, and Cybernetics*, 1997, vol. 3, pp. 2269–2274.

[90] B. Skyrms, *Choice and Chance: An Introduction to Inductive Logic*. Belmont, CA: Wadsworth, 1999.

[91] L. B. Solum, "Legal personhood for artificial intelligences," *North Carolina Law Rev.*, vol. 70, no. 1, pp. 1231–1287, 1992.

[92] R. Sparrow, "The march of the robot dogs," *Ethics Inform. Technol.*, vol. 4, no. 4, pp. 305–318, 2002.

[93] R. Sparrow and L. Sparrow, "In the hands of machines? The future of aged care" *Minds Mach.*, vol. 16, no. 2, pp. 141–161, 2006.

[94] C. M. Stanton, P. H. Kahn, Jr., R. L. Severson, J. H. Ruckert, and B. T. Gill, "Robotic animals might aid in the social development of children with autism," in *Proc. 3rd ACM/IEEE Int. Conf. Human Robot Interaction*, 2008, pp. 271–278.

[95] J. P. Sullins, "When is a robot a moral agent?" *Int. Rev. Inform. Ethics*, vol. 6, no. 12, pp. 23–30, 2006.

[96] Surgeon General's Office. (2006, Nov. 17). Mental health advisory team (MHAT) IV operation Iraqi freedom 05-07. Final Report [Online]. Available: http://www.armymedicine.army.mil/reports/mhat/mhat_iv/MHAT_IV_Report_17NOV06.pdf

[97] A. Tapus, C. Tapus, and M. J. Matarić, "Music therapist robot for individuals with cognitive impairments," in *Proc. 4th ACM/IEEE Int. Conf. Human Robot Interaction*, 2009, pp. 297–298.

[98] R. Tonkens, "A challenge for machine ethics," *Minds Mach.*, vol. 19, no. 3, pp. 421–438, 2009.

[99] US Department of Defense, Office of the Assistant Secretary of Defense (Public Affairs) News Transcript. (2007, Dec. 18). DoD Press Briefing with Mr. Weatherington from the Pentagon Briefing Room, Arlington, VA [Online]. Available: http://www.defense.gov/Transcripts/Transcript.aspx?TranscriptID=4108

[100] G. Veruggio. (2005). The birth of roboethics. presented at Workshop on Robo-Ethics, IEEE Int. Conf. Robotics and Automation (ICRA) [Online]. Available: http://www.roboethics.org/icra2005/veruggio.pdf

[101] G. Veruggio, "The EURON roboethics roadmap," in *Proc. 2006 6th IEEE-RAS Int. Conf. Humanoid Robots*, 2006, pp. 612–617.

[102] G. Veruggio and F. Operto, "Roboethics: A bottom-up interdisciplinary discourse in the field of applied ethics in robotics," *Int. Rev. Inform. Ethics*, vol. 6, no. 006, pp. 2–8, 2006.

[103] K. Wada and T. Shibata, "Living with seal robots its sociopsychological and physiological influences on the elderly at a care house," *IEEE Trans. Robot.*, vol. 23, no. 5, pp. 972–980, 2007.

[104] K. Wada, T. Shibata, T. Musha, and S. Kimura, "Robot therapy for elders affected by dementia," *IEEE Eng. Med. Biol.*, vol. 27, no. 4, pp. 53–60, 2008.

[105] K. Wada, T. Shibata, T. Saito, K. Sakamato, and K. Tanie, "Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged," in *Proc. 2005 IEEE Int. Conf. Robotics and Automation (ICRA)*, 2005, pp. 2785–2790.

[106] M. Walzer, *Just and Unjust Wars*, 3rd ed. New York: Basic Books, 2000.

[107] V. Wiegel and J. van den Berg, "Combining moral theory, modal logic and MAS to create well-behaving artificial agents," *Int. J. Soc. Robot.*, vol. 1, no. 3, pp. 233–242, 2009.

[108] K. Dautenhahn, C. L. Nehaniv, M. L. Walters, B. Robins, H. Kose-Bagci, N. A.f Mirza, and M. Blow, "KASPAR—A minimally expressive humanoid robot for human-robot interaction research," *Appl. Bionics Biomech.*, vol. 6, no. 3, pp. 369–397, 2009.

Pawel Łichocki, Laboratory of Intelligent Systems, Ecole Polytechnique Federale de Lausanne, 1015 Lausanne, Switzerland. E-mail: pawel.lichocki@epfl.ch.

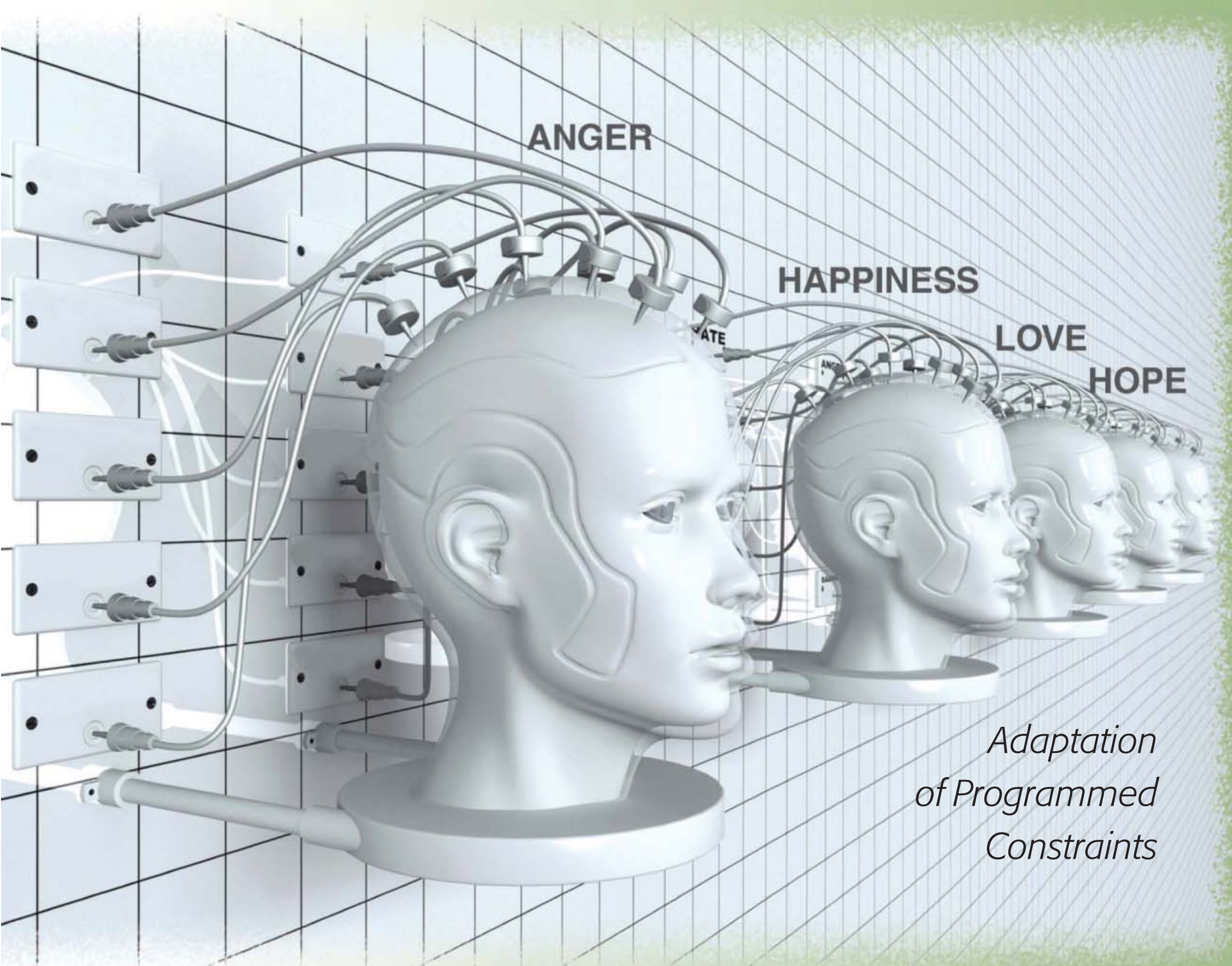
Peter H. Kahn, Jr., Department of Psychology, University of Washington, Seattle, WA 98195-1525. E-mail: pkahn@u.washington.edu.

Aude Billard, Learning Algorithms and Systems Laboratory, Ecole Polytechnique Federale de Lausanne, 1015 Lausanne, Switzerland. E-mail: aude.billard@epfl.ch.

Robotics & Automation Magazine IEEE | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

Incremental Machine Ethics

By Thomas M. Powers



Adaptation of Programmed Constraints

Approaches to programming ethical behavior for computer systems face challenges that are both technical and philosophical in nature. In response, an incrementalist account of machine ethics is developed: a successive adaptation of programmed constraints to new, morally relevant abilities in computers. This approach allows progress under conditions of limited knowledge in both ethics and computer systems engineering and suggests reasons that we can circumvent broader philosophical questions about computer intelligence and autonomy.

Digital Object Identifier 10.1109/MRA.2010.940152
Date of publication: 14 April 2011

1070-9932/11/$26.00©2011 IEEE

IEEE Robotics & Automation Magazine
Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

Incrementalism is the view that progress toward a goal is made in a stepwise fashion; it is thought to be applicable, especially, in circumstances in which it is difficult to know at the onset what will be the proper and efficient means that will allow the goal to be reached [11]. Uncertainty about the means typically owes to the complexity of the problem to be solved, and incrementalism becomes more attractive as the need for some progress toward the goal (however, piecemeal) becomes more pressing. Machine ethics, in the sense in which I will here discuss it, is the goal of properly constraining computer-controlled machines (robots and other multifunction computer systems), where the grounds for the constraints are the ethical reasons. Since machine ethics as an academic discipline is in its infancy, there are basic philosophical questions surrounding its plausibility—ones that concern the nature of moral agency and responsibility for non-human actors—it is perhaps unwise to wait for a complete philosophical account of its objectives and methods. For instance, in the definition of machine ethics mentioned earlier, we could easily get sidetracked by questions over whether the ethical reasons need to be unanimously endorsed by humans or whether they would count as reasons for the computer. Basic issues in philosophy, as any historian of the discipline will attest, take a long time even to formulate correctly, and some of them may never be resolved.

Incrementalism is the view that progress toward a goal is made in a stepwise fashion.

When these two notions are combined, incrementalism in machine ethics becomes a practical proposal about how to simultaneously engineer and provide an ethical sanction for the kinds of information technologies that are taking over the many activities—once performed solely by humans—that are generally assumed to have moral relevance. In fact, technological societies have already traveled some distance down the path of replacing human action with technology. We now have technologies that assist us to fight wars, keep public order, monitor air and water quality, provide medical care, execute financial transactions, distribute electrical power, and so on. These activities have moral relevance because human and animal lives and welfare depend on them. Many people living in contemporary societies are dependent on machines for their well being, and they will likely become more dependent as machines gain in their functionality and are deployed in further domains. So while we are already getting the machines (like it or not), we desperately need the ethics that ought to accompany them. This is why I call machine ethics a pressing problem.

Versions of incrementalism have already been developed in academic studies of social choice under conditions of bounded rationality [18], of the federal budgeting process [21], and in international relations [1]. Since the 1960s, incrementalism has been at the center of many debates in the social sciences. Incrementalists, in general, share a criticism of the *synoptic* or rationalist-comprehensive view of public administration, political science, and social choice. The starting point for this criticism is the recognition that, in trying to solve complex social problems, humans suffer from cognitive limitations and the decentralization and wide distribution of mechanisms for decision making [6]. The rationalist expectation that policy scientists must start out already knowing how to proceed often leads to wasted theorizing or a paralysis of indecision. The central study with which incrementalism is most often identified came from former Research and Development Corporation researcher Charles Lindblom's work on the social science of public administration that began with his article "The Science of 'Muddling Through'" [12] and continued in series of widely discussed publications, chief among them being *The Intelligence of Democracy* [13]. To account for partisanship in public policy decision making, Lindblom developed with the philosopher David Braybrooke the theory of disjointed incrementalism [4]. Lindblom's classic account included the notion of adaptive incremental adjustment to respond to a prior decision in the strategic decision space [11]. Though incrementalism has met with resistance from rationalist economists and political scientists [3], [15], Lindblom's account of muddling through has become one of the most widely read pieces in the social sciences [16].

Incrementalism has now emerged in implicit forms in the recent literature on machine ethics. Wallach and Allen [20] profess an interest in the incremental steps arising from present technologies that suggest a need for ethical decision-making capabilities and explore the prospects for a bottom-up approach to building an artificial moral agent (AMA) that has many of the features of an incrementalist approach. The notion of adding ethical constraints as a machine takes on new functions is connected to Johnson and Powers' account of the role responsibility of computers as surrogate agents [9]. But the roboticist Ron Arkin has come closest to advocating an explicitly incrementalist position. Much of Arkin's work concerns lethal autonomous robots in the context of warfare. In describing the step-by-step development of a fieldable ethical system for such robots, Arkin writes that "we can initially represent a small set of forbidden and obligated constraints and test the overall system without the necessity of a fully complete set of representational constraints" [2]. He thinks that, given the impending widespread introduction of lethal warfare robots, "baby steps are better than no steps toward enforcing ethical behavior in autonomous system warfare."

Abstracting from these studies in the machine ethics literature, we can explore in greater depth what it means to be committed to incrementalism in machine ethics, both as a description of the (very young) practice of applying ethical constraints to machines and as a normative model for the development of machine ethics for highly integrated and safety critical computer systems, such as those for warfare, air traffic control, and public health. The

Robotics & Automation MAGAZINE IEEE | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

supposed benefit of thinking about progress toward a goal from an incrementalist perspective is that doing so will avoid the often-paralyzing search for a final theory when circumstances are not ripe for understanding what that theory might look like. The contrasting synoptic view of machine ethics suggests that we could conceive of a machine ethic in its totality, prior to knowing what capabilities the machine will have or the time frame during which they will be developed. Incrementalists, on the other hand, need not start out with the assumption that they know where they are going; they just want to decide (given the circumstances): what's the next step? I start off, then, with several basic suppositions. There is a pressing need for progress in machine ethics, and we have no final theory of ethics, nor any good idea of how the steps of a machine ethical theory might develop.

In what follows, I will try to explain and evaluate incrementalism for machine ethics. Borrowing a term from decision theorists [11], I will sketch a particular version that I call *adaptive incrementalism in machine ethics* (AIME), and describe how it might address both practical and philosophical problems that have already become apparent in the literature. In the final section, I will look at the potential criticisms of AIME, especially the suggestion that a machine ethic that is developed in a piecemeal fashion must remain incomplete. (This criticism can be summed up thus: "you can't get there from here—incrementally or otherwise") I will then introduce an argument from limited behaviorism that circumvents the criticism that no machine can be said to act ethically without our first having established that it is a free moral agent. The argument from limited behaviorism, I will urge, does not reduce to a general behaviorism for human ethics.

Limited behaviorism does, however, provide a response to an objection to machine ethics that is based on an objection to machine intelligence—an objection made famous by John Searle's Chinese Room argument. The Searlean objection to machine ethics is that a machine could never become ethical because it cannot develop—either incrementally or otherwise—to a point at which it would become be conscious. What the objections to machine intelligence and machine ethics share are doubts about a computer's ability to have intentional states (including states like "intending to act" and "preferring X over Y"). Machine ethics must also face the longstanding supposition that having and acting because of intentional states are necessary conditions for something being a moral agent. While the objection to machine intelligence is too complex to address in the space available, the objection to machine ethics must be met, for if it is correct, then it will be just as senseless to talk of machine ethics as to talk of machine social psychology.

Adaptive Incrementalism

As a model to develop machine ethics, incrementalism operates on two levels: as adaptations to the commands or program of the machine (software level) that act as ethical constraints and as additions to the capabilities of the hardware/software (system level) that trigger new ethical constraints. Let us begin with the latter first.

Candidates for the kinds of machines that will need ethical constraints have a hierarchy of system capabilities. While we need not order these capabilities, we generally know that the stored program architecture, for instance, is more basic than the ability to implement a voice recognition program, or to play a game of chess, or launch a rocket. Some approximate division can be made between the basic system capabilities and the morally relevant ones. As a start, let us say that

1) The computer system gains morally relevant capabilities as soon as some human being could be made worse off by the designed action of the system.

> **Incrementalists share a criticism of the *synoptic* or rationalist-comprehensive view of public administration, political science, and social choice.**

This formulation eliminates the concern for accidental harms that might come, for instance, from a CPU falling on someone's foot or causing an electrical fire because of a faulty wire. Still, it casts a very wide net for morally relevant capabilities and opens up the door for disputes about moral theory. Surely, we will have to look closely at the different ways in which someone could be made worse off, and whether these kinds of harms count as morally relevant. As a first refinement, let us adopt Pareto's criterion of efficiency and say that

2) The computer system gains morally relevant capabilities as soon as some human being could be made worse off (and no one better off) by the designed action of the system.

This is a narrower formulation, but it will not be completely uncontroversial. We could have competing triggers for moral relevance; utilitarians would insist on finding the point at which the designed actions promise to produce more disutility than utility; rights theorists would want to use a list of basic rights; and so on. The point here is not that the issue of the trigger can reintroduce into machine ethics all of the same disputes we find in moral theory. Rather, it is that there will be some additional system capability that is seen as violating Paretian efficiency—perhaps because it violates someone's rights and thereby diminishes total social utility—and the addition of this capability will trigger consideration of the first level of machine ethical incrementalism: the software level.

Changes to the machine software to provide an ethical constraint might also bring with them the possibility of differing prescriptions according to the favored moral theory. Here though we can see that the response embodied by the programmed ethical constraint must have one of the three abstract descriptions:

3) Allow the machine to act on the capability unchanged.
4) Meliorate the way in which the human being might be made worse off.
5) Disallow the capability entirely.

As an example, consider the addition of the ability to transfer files to and from another networked computer through the standard file transfer protocol (FTP). One ethical constraint would be to add a capability to lock certain files in the target computer, thereby disallowing reading or writing to the files. Another constraint would be the introduction of password-protected access such as secure file transfer protocol (SFTP).

Candidates for the kinds of machines that will need ethical constraints have a hierarchy of system capabilities.

The previous example is almost trivial in its ethical import. Consider, however, the additional capabilities that networked computers gained when suddenly they could attach viruses to e-mail or hack into a target computer and replace system files. The addition of these capabilities already has called forth adaptations: firewalls and antivirus software. This is an example of AIME, in practice, and one that comes fairly early in the young history of machine ethics. It is interesting that such changes were broadly adopted without large public debates about the correct moral theory to use. Viruses were deemed bad, no matter what moral theory one held.

The serial implementation of constraints of the sort in 3)–5) would constitute the development of an adaptive machine ethic. At another level of abstraction, the list of ethical constraints for some machine at a certain point in time might read as follows:

6) Protect the privacy of clients.
7) Protect the property rights of clients.
8) Maintain the health of bystanders.

At a more fine-grained level, the designers of the ethical constraints might describe 6) as prevent unauthorized access to files containing medical data of client and similarly for each ethical constraint we could have more clearly operationalized commands for the machine to follow. This feature of differing descriptions of constraints, based on the level of abstraction, is also apparent in human ethics. People never directly act on the constraint don't be bad. Rather, someone's action might be governed by the constraint don't cause needless harm to people or even more concretely, don't strike dan now.

An important feature of AIME is that there is no a priori list of ethical constraints for a machine; each constraint is developed because of, or in response to, an additional capability. This means that there are no necessary components of an adaptive incremental ethic. Also, there is no limitation, in principle, on the entities that will gain moral status and thus come under consideration when thinking about who (or what) is potentially made worse off by the computer's new capability. In the formulation given earlier, I included only human beings in 1) and 2), but a machine ethic could develop to include animals, ecosystems, future generations of humans, etc.

To most ethicists, the potential conflict of values in 6)–8) (privacy, property, and health) will be readily apparent. If there is a software adaption that presents a choice between these three values, then there will be no theory-independent way of ranking the outcomes. The problem of conflicting values is not so acute in situations where one and the same machine receives, serially, the adaptations in 6)–8); this merely represents a pluralism of values. Still, what happens when we must choose only one adaptation—when, for instance, we must choose to protect someone's health at the expense of someone else's informational privacy?

Lindblom's account of muddling through addresses this very problem, albeit in the context of the conflicting values that every public administrator must consider. He states the problem thus:

> How does one state even to himself the relative importance of these partially conflicting values? A simple ranking of them is not enough; one needs ideally to know how much of one value is worth sacrificing for some of another value. [16]

and provides the following answer:

> The value problem is . . . always a problem of adjustment at a margin . . . . That one value is preferred to another in one decision situation does not mean that it will be preferred in another decision situation in which it can be had only at great sacrifice of another value [12].

The key to the incrementalist's sanguine acceptance of value conflict is that he gives up on a rational-comprehensive account of the system as a whole. Lindblom admits that "[e]xcept for roughly and vaguely, I know of no way to describe—or even to understand—what my relative evaluations are for, say, freedom and security, speed and accuracy in governmental decisions" [12]. That very same uncertainty about relative evaluations might hold if we replace governmental decisions in Lindblom's statement with information technology decisions. For AIME, this means that ethically adapted robots and computer systems will betray the multiple ethical perspectives of their designers, and indeed, this multiplicity will be unattractive to proponents of a rational-comprehensive view of machine ethics.

For the incrementalist, though, appeal to a rational-comprehensive view is not useful for actual ethical decisions, because in the context of decision making, it is not a view available to him or her. Computer systems, like other technologies, often evolve over time frames that encompass different users and designers (indeed, sometimes, even different teams of designers); they are affected by laws and regulatory policy, markets, available infrastructure, and changes thereto [8]. No designer has the option to select one instantiation of the system to last forever. All changes are temporary in principle.

Following Lindblom's account of incrementalism, we would describe each choice to adopt an ethical constraint as an instance of a successive limited comparison. The relevant question for this comparison is, does the designer prefer the new system—with both its additional, morally relevant ability and its programmed ethical constraint—to the system as it was? Of course, there will always be the possibility that the designer's preference will diverge from some users' preferences, but the fact is that a choice must be made. We should not assume that a designer's preferences are independent of those of the users nor that they are equivalent. In any event, the user is not in a position (typically) to make design decisions, and the designer's knowledge is limited by many factors.

AIME thus represents a description and a normative proposal for design of machines from engineering and ethical perspectives. We may engineer certain capabilities into a machine and find, at a later date, that there was a better way to achieve the same results. Likewise, we may put in place a certain ethical constraint for a machine, to adapt it to society, given its increased functionality. At some later date, we may find that this is the wrong constraint or that there is a more precise way to constrain the machine—allowing it to do more or do less than we originally allowed it to do. Incremental adaption therefore suggests an ongoing process to address new machine capabilities and to reevaluate old constraints in light of the new capabilities.

Incrementalism, as applied to the problem of machine ethics, should have the same advantages as those promised for the original incrementalism—the science of muddling through in public policy. In his study of incrementalism for policy making, Hayes identifies five virtues of incrementalism (summarized here):

1) facilitates action where the rational ideal is paralyzed
2) reduces the costs of analysis by providing a defensible basis for confining attention to some alternatives over others
3) facilitates learning from mistakes
4) facilitates majority building my minimizing disruption to established practices
5) the failure of any given step to solve a particular problem often makes the best case for taking the next step.

Advocates of incrementalism claim to have confirmed these characteristics in empirical studies in several areas of the social sciences, while critics have countered with other studies. As Knott et al. point out, "[t]he concept of incremental adaption entered the social sciences literature because empirical observations of behavior did not fit with a fully rational approach to decision making" [11]. So even if some studies did validate the rational-comprehensive model, the advocates of incrementalism were able to argue that that model works in far too few circumstances to be adopted. In addition, they were able to show that the logic of incrementalism under conditions of cognitive limitations (bounded rationality) promised more success for decisions under those circumstances [14]. Much of the force behind incrementalism as a movement in the social sciences thus depended on both a priori and a posteriori arguments for it, as its advocates were able to shift adeptly between descriptive and normative conceptions of the model.

Criticisms of AIME

It is safe to say that theorists and practitioners of machine ethics—a diverse group consisting mostly of philosophers, ethicists, computer engineers and programmers, and artificial intelligence (AI) enthusiasts—are predisposed to rationalist accounts. Hence, it is unlikely that a muddling through approach to machine ethics would be adopted tout court. We will now turn to consider some of the challenges for the AIME version of incrementalism.

Since adaptive ethical constraints in AIME are triggered by system-level changes to machine capabilities, a question of scope will arise even before the first constraint is operationalized. The question is, are all capabilities of machines relevant? My earlier refinement of the definition of a trigger for morally relevant system changes attempted to forestall this worry. It is likely, though, that some designers will initially be stumped and may consider too many capabilities as relevant. The other extreme, of course, is to see the trigger as operating hardly at all. This latter outcome, I believe, is our current situation with computer system development. Without ethicists participating in design, almost all system-level abilities are taken to be morally irrelevant. (Arkin, a roboticist, is one of the few who takes moral relevance seriously) To overcome this potential problem with AIME, designers and ethicists will have to work jointly toward a proper moral sensitivity for the computer systems they are designing. This will require a good deal of imagination—an appreciation for plausible what if? scenarios—and also timely feedback about their systems in the testing and initial deployment phases.

Machine ethics must also face the longstanding supposition that having and acting because of intentional states are necessary conditions for something being a moral agent.

Conceivably, the ethicists and designers may fail to develop the same level of sensitivity, and this could lead to disagreements and conflicts. The second major criticism, then, questions whether it is really plausible in AIME to expect people from different backgrounds and with disparate objectives to come together to make steady progress. If this isn't possible, then AIME will suffer the same fate as the incrementalists found in rationalist-comprehensive practice: paralysis. One way to allay the worry about internecine disputes among teams of designer ethicists is to impose on the team the same set of institutional interests. For private sector projects, for instance, a company might tie compensation or promotion to the total performance of the robot or computer system. That is, they would be wise to consider a successful system as one that meets simultaneously the criteria of high functioning and ethical propriety.

Making AIME a corporate benchmark may not entirely solve the problem, however. Some computer systems could be of a scale that they would pit the interests of one corporation against another or even one nation against another. This will most likely be the case with warfare robots. For instance, the primary ethical constraint for some military commanders will be to minimize friendly fire casualties. Of course, the system may be programmed also to respect the laws of war, minimize collateral damage, protect noncombatants, and observe proportionality of destructive force. As is clear from Arkin's prototype implementation [2], not every objective can be pursued at once. How these objectives are weighted may vary between nations and even between computer systems.

One feature of AIME that may make some (but certainly not all) ethicists uncomfortable is the lack of theoretical unity of ethical constraints that are developed over time. This concern also affects the definition of the system trigger for ethical constraint. But the heterogeneity of these machine ethics may go deeper. So far, we have spoken merely of ethical constraints, but (allowing for developments in software) it will likely happen that complex machines in the far future will also need ethical rules, perfect and imperfect duties (in the Kantian sense), and maybe even a conception of virtues. This eventuality is not troubling for the incrementalist, since the motivating assumption is that machine ethics starts out by doing what it can do and not worrying about what it cannot do in imparting ethics to machines. It is likely that, over time, the programmed ethical constraints of a particular system will be superseded by future constraints. It is even possible that entire schemes of machine ethics that were once considered successful will be found at some point to be inferior to new schemes. This mimics one of the most interesting aspects of the history of the electronic digital computer. Hardware configurations, storage media, input/output, and programming languages undergo revolutions of sorts (though not all simultaneously) when improvements are developed. We should expect the same for machine ethics.

AIME thus represents a description and a normative proposal for design of machines from engineering and ethical perspectives.

Ethics, Consciousness, and Agency

The very notion that machines could take on ethical abilities faces philosophical challenges from many quarters, most of which we cannot consider here. Comparing machines to typical (human) moral agents, philosophers have insisted that machines lack free will, consciousness, and morally relevant emotions such as regret, empathy, and shame. Perhaps the most succinct complaint against attributing excessive abilities to computers comes in John Searle's attack on the notion that computers can think [17]. This suggests an objection to machine ethics by means of a rather obvious extension of Searle's argument against computer thought.

Searle denies that computers have anything more than syntactical abilities in operating according to their programs. He thinks that semantic ability—something only minds have, as far as we know—is the key to conscious understanding. In addition to lacking semantic ability, computers lack the ability to have their own intentional states, on his view. Certainly, they can represent intentional entities—for instance, with sentences of a natural language displayed on a screen. And they can follow syntactic rules to display new sentences, as his Chinese room argument suggests. This is a mere simulation of thinking, according to Searle; it is not the real thing. Simulation, in the sense of supplying answers to queries, isn't sufficient for thinking, understanding, or consciousness, according to Searle. The case against machine intelligence, for Searle, is simply open-and-shut.

When it comes to the free will of a computer, here Searle is somewhat less emphatic. He concedes that if "somebody built a robot that we became convinced had consciousness, in the same sense that we do, then it would at least be an open question whether or not that robot had freedom of the will" [17]. Nonetheless, he is quite sure that this will not happen, as he thinks the computer would have to have the abilities of a human brain in order for it to have consciousness and intentional states of its own.

Why does Searle's syntax/semantics argument against AI not doom AIME? First, we must realize that consciousness in machines is not a necessary condition of their ethical behavior. It may well be a necessary condition for a machine to be self-aware, and hence aware of its intentional states—should it have them—as being the free cause of its actions. But at this point, we should acknowledge a point that Searle takes to be in his favor, but that actually cuts against his arguments. We poorly understand the physical and neurological bases of consciousness and intentionality that the emphatic argument against machine intelligence overreaches what we know. When we do come to understand the brain better, it may well turn out that intentionality and consciousness is possible for computers. Searle's syntax/semantics argument is just too simple to rule out that possibility.

Second, we might understand intentional states in computers as merely those states that are about, or represent, or are directed at states of the world through the models that they use. As the computer scientist Brian Cantwell Smith wrote, "there is no computation without representation!" [19]. When I look at a radar animation, from a computer system, of storms gathering over Delaware, those images and the representations that generated them are about the weather outside. This weaker sense of intentionality allows that computers have intentional states—in fact, they must have them insofar as they have representations that connect to models of the world. Computers do not (now) generate their own intentional states, independent of the representations that we program them to have. But as Smith points out, we put representations in computers—representations that serve as a partial model of the world—when we design them. The having of representations about the world is sufficient for computers' having intentional states in the weak sense.

If we attribute weak intentional states to computer systems and suppose that AIME is successful in constraining the behavior of the system—for ethical reasons—we have the essentials of a system that could be seen to behave ethically in a world that it represents in it programs. Having arrived at the heart of the issue, this behavioral definition of machine ethics, I will argue, is a genuine beginning to machine ethics.

To see how this is so, consider what I will call the causal–rational tradition in ethical theory from modern and contemporary philosophy—one exemplified in the works of Donald Davidson [5]. According to this tradition, when we act ethically, we act for (or with) a certain reason, but not any old action for a reason will suffice. An outwardly right action done for the wrong reason is not an ethical action, but neither is it necessarily unethical. For instance, paying a debt because you fear retribution or saving a child's life because you thought it was your son (but you were mistaken) are taken to be ethically neutral on this view. Right act, wrong reason.

If not all (outwardly appearing) right acts are ethical acts, and the ethical acts are determined by the reasons that caused them, what determines what the right acts are? This has been a challenge that the causal–rational tradition has yet to answer. Let us sketch a way around this problem.

Suppose some (not inconceivable) future in which all of humanity converges on one theory of ethics. Call it theory T. We settle all disputes about which things have moral status, what we owe to them, etc. To perform all of the right actions, we know all of the self and other regarding obligations, we know the correct virtues and how to practice them, we know which preferences count, and we know how much they count. That is, we know all of the ethical reasons for all of the right acts, and we humans are now able to do the right acts for the ethical reasons. Consider building a machine that fulfilled all of the right acts of theory T because it was programmed through AIME. (By this time, our moral trigger would have become much more sophisticated.) For instance, if T consists of obligations and permissions, the machine acts on all of the obligations in the appropriate contexts and never commits an impermissible act. What would we have to say then about the ethics of that machine?

The opponent of AIME might claim that the machine didn't do the right acts (as determined by our complete theory T) for the right reasons, and in this, she would be correct, but only because the machine didn't do the acts for any reasons at all. Machines, the opponent insists, aren't capable of acting for a reason. The proponent of AIME should, I think, accept this criticism. Moreover, this very fine result—that the machine does exactly what T prescribes—is all we could want anyway, given our current state of knowledge about computing machines and ethics. This is a result that I call *limited behaviorism*: a machine behaves ethically in doing all and only the right acts. This kind of behaviorism does not offer a replacement for a comprehensive ethical theory for humans, but it allows that, in principle, we may develop a machine that performs all of the right acts (and none of the wrong ones) that we would expect of any human operating according to theory T. That is, limited behaviorism is not reductive on the issue of ethical reasons for humans; a human acting for ethical reasons is still defined separately from the right acts that the human performs. But limited behaviorism does accept the equivalence of right acts, whether they issue from a machine or a human. It also finds that, since there are no machines that act for ethical reasons, there will be no practical difference between a right-acting machine and an ethical machine. In the event that machines develop to the point where, one day, they can have reasons for acting, they too would have to distinguish their (outwardly) right acts from their ethical acts.

One feature of AIME that may make some ethicists uncomfortable is the lack of theoretical unity of ethical constraints that are developed over time.

Finally, the AIME opponent might raise a different kind of objection: that we're nowhere close to figuring out what theory T is, so if we go about designing and programming a machine with our current imperfect state of moral knowledge, we are bound to impart some mistakes to the machine's ethical build. Compare now our current attitude toward the inculcation and practice of ethics for humans, in this (quite imperfect) state of moral knowledge in which we find ourselves. Most of us are quite sure—on both theoretical and practical levels—that we aren't close to discovering theory T. Is that fact a reason for not teaching ethics

Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

to our children? Is that fact a reason to give up, ourselves, on trying to do what is right? I think the answer to these questions is clearly no.

The propriety of muddling through when it comes to machine ethics is reinforced by an analogy to the theory of moral development in the cognitive/structural account of moral education that was introduced by the Swiss psychologist Jean Piaget and is now widely associated with Lawrence Kohlberg's theory of the stages of moral development [10]. Wallach and Allen address Kohlberg's theory explicitly as a useful way of thinking about machine ethics [20]. The basic idea in Kohlberg's work is that no child is born with its full complement of moral abilities, and some never develop to the highest stage of moral development, but there is progress in teaching a child to reach a higher stage. Indeed, after Kohlberg's theory became well known, many scholars came to doubt whether he had correctly described and ordered the stages. But as far as I know, no serious scholar ever proposed that parents and teachers cease moral education of children until the theorists could come up with one comprehensive account of both moral development and moral knowledge.

In the spirit of Kohlberg, roboticists might set themselves the task of constantly refining the ethical abilities of their machines. Yet, it is unclear that there is an ultimate stage of ethical behavior to be reached. Ethical performance may be relative to the kinds of tasks to be performed, and the moral complexity of the environment in which the machine operates. If this is so, the requirements of machine ethics may always get more demanding. Human parents might hope to teach their children to reflect on the nature of that ultimate stage—something that is likely to be exceedingly difficult to program into a machine.

This analogy between parenting and programming returns us to the main distinction in the beginning of this article: that incrementalism is a means for the goal of machine ethics. Incrementalism does not purport to list the stages in the development of machine ethics, and our current poor state of moral knowledge does not provide a clear-enough picture of what a right acting machine will look like, even on the assumption that it can be built. Thus, ethicists and designers find themselves in a situation like that of new parents. They have responsibility for a new, developing being, but aren't sure how best to effect its moral education. They themselves lack complete moral knowledge—they await (but despair of never discovering) a perfect theory T of morality. They would like to be able to impart practical reasoning to this new being but realize that this is a job for later years, and—first things first—they have to get the child simply to behave. They are faced with a daunting task, but one that is pressing due to the likelihood that doing it imperfectly, by muddling through, is bound to be better than not doing it at all.

Acknowledgments

This work was supported in part by the U.S. National Science Foundation under grant EPS-081425.

References

[1] G. Allison, *Essence of Decision: Explaining the Cuban Missile Crisis*. Boston, MA: Little Brown, 1971.

[2] R. C. Arkin, *Governing Lethal Behavior in Autonomous Robots*. Boca Raton, FL: CRC Press, 2009.

[3] K. Arrow, "A strategy of decision: Policy evaluation as a social process by David Braybrooke; Charles E. Lindblom," *Polit. Sci. Q.*, vol. 79, no. 4, pp. 584–588, 1964.

[4] D. Braybrooke and C. E. Lindblom, *A Strategy of Decision*. New York, NY: The Free Press, 1963.

[5] D. Davidson, *Essays on Actions and Events*. New York, NY: Oxford Univ. Press, 2001.

[6] J. Forester, "Bounded rationality and the politics of muddling through," *Public Admin. Rev.*, vol. 44, no. 1, pp. 23–31, 1984.

[7] M. Hayes, *The Limits of Policy Change: Incrementalism, Worldview, and the Rule of Law*. Washington, DC: Georgetown Univ. Press, 2001.

[8] T. Hughes, "The evolution of large technological systems," in *The Social Construction of Technological Systems*, W. Bijker, T. Hughes, and T. Pinch, Eds. Cambridge, MA: MIT Press, 1989.

[9] D. Johnson and T. M. Powers, "Computers as surrogate agents," *Information Technology and Moral Philosophy*, M. J. Van den Hoven and J. Weckert, Eds. New York: Cambridge Univ. Press, 2008.

[10] L. Kohlberg, *Essays on Moral Development,* vol. I: *The Philosophy of Moral Development*. San Francisco, CA: Harper & Row, 1981.

[11] J. Knott, G. Miller, and J. Verkuilen, "Adaptive incrementalism and complexity: Experiments with two-person cooperative signaling games," *J. Public Admin. Res. Theory*, vol. 13, no. 3, pp. 341–36, 2003.

[12] C. Lindblom, "The science of "muddling through," *Public Admin. Rev.*, vol. 19, no. 2, pp. 79–88, 1959.

[13] C. Lindblom, *The Intelligence of Democracy*. New York: The Free Press, 1965.

[14] C. Lindblom, "Still muddling, not yet through," *Public Admin. Rev.*, vol. 39, no. 6, pp. 517–526, 1979.

[15] I. Lustick, "Explaining the variable utility of disjointed incrementalism: Four propositions," *Amer. Polit. Sci. Rev.*, vol. 74, no. 2, pp. 342–353, 1980.

[16] R. Premfors, "Review article: Charles Lindblom and Aaron Wildavsky," *Brit. J. Polit. Sci.*, vol. 11, no. 2, pp. 201–225, 1981.

[17] J. Searle, *Minds, Brains, and Science*. Cambridge, MA: Harvard Univ. Press, 1984.

[18] H. S. Herbert, "A behavioral model of rational choice," in *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. New York: Wiley, 1957.

[19] B. Smith, "Limits of correctness in computers," *ACM SIGCAS Comput. Soc.*, vol. 14, no. 4, pp. 18–26, 1985.

[20] W. Wallach and C. Allen, *Moral Machines: Teaching Robots Right from Wrong*. New York: Oxford Univ. Press, 2009.

[21] A. Wildavsky, *The Politics of the Budgetary Process*. Boston, MA: Little, Brown, 1964.

Thomas M. Powers, University of Delaware, Newark, DE, 19716, USA. tpowers@udel.edu.

The Robot DustCart

By Pericle Salvini, Giancarlo Teti, Enza Spadoni, Cecilia Laschi, Barbara Mazzolai, and Paolo Dario

IMAGE COURTESY OF PERICLE SALVINI

Focus on Social and Legal Challenges

Peccioli, a small medieval town in Italy, became one of the first places in the world where a robot was used (not demonstrated) to carry out a public service in the urban environment (from 15 June 2010 to 7 August 2010). Thirty-five real users *accepted* to trash their domestic waste using the robot DustCart, a mobile robot designed to collect, transport, and discharge rubbish bags in complete autonomy. During the testing period, the robot safely traveled along the public streets of Peccioli, carrying out its daily service and sharing the urban environment with the passers-by, bicycles, and cars, without causing any problems. Drawing on this unique event, in which the authors also participated, the article addresses some of the implications originating from the actual deployment of autonomous mobile robots in urban areas. Our reflections will gravitate around two major issues: legal regulations and social acceptance. More specifically, we will report on the legal solutions adopted for deploying DustCart in the streets of Peccioli and the activities carried out to increase the social acceptance of the robot.

Till today, the deployment of autonomous mobile robots in urban environments has been the talk of science fiction. A memorable example is a short story and the movie based on it, *I Robot* [1], where the robots carry out various kinds of services in human-inhabited settings. In a particular scene, humanoid robots are walking down the street, shoulder to shoulder with human beings. This is an exemplary case of coexistence between human beings and robots. In this article, we recount a similar story, but this time it is based on real facts: that of a service robot called DustCart, which was used for more than a month in a small Italian town to collect rubbish bags and then transport them to a discharge site. The robot, which was designed and developed within the framework of the European Union (EU) project DustBot [1], traveled on public roads in complete autonomy, interacting with people and cars and coexisting in the urban life of Peccioli. As far as we know, there are no references in literature to service robots being deployed in an urban environment or for such a lengthy period of time.

The objective of this article is to report on the testing of the robot DustCart in Peccioli and to point out some of the ethical, legal, and social implications that emerged before and during the test period.

Digital Object Identifier 10.1109/MRA.2010.940153
Date of publication: 14 April 2011

1070-9932/11/$26.00©2011 IEEE

IEEE Robotics & Automation Magazine
Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

Ethical Triaging

The different aspects that make autonomous mobile robots such an ethically sensitive topic can be illustrated by drawing on the triage method. The triaging method has been previously used in ethical research in [2] for investigating the ethical implications that arise from the research and deployment of brain–computer interfaces. It has also been used in the framework of the EU-funded project ETHICBOTS [3], where it was applied to select ethically sensitive items, namely, technologies or applications that were deemed worthy of ethical investigation. The triaging method consists of analyzing a given technology according to the following two criteria: imminence and social pervasiveness.

Peccioli became one of the first places in the world where a robot was used (not demonstrated) to carry out a public service in the urban environment.

As to imminence, it measures the level of maturity and availability of enabling technologies. In our investigation, the enabling technologies refer to the robots' ability to move autonomously in partially unstructured environments, which include navigation, obstacle avoidance, environmental perception, and self-localization. Recently, these technologies have progressed substantially. A significant illustration of the advancements achieved in the field of autonomous mobile robots is provided by the Defense Advanced Research Project Agency Grand Challenge. The 2007 edition of the challenge included a competition among driverless cars that moved autonomously in an urban environment [4]. Other relevant illustrations of technological maturity in the field of autonomous navigation can be found in a group of projects funded in the sixth framework programme of EU. We refer to the research project DustBot [5], which will be discussed in greater detail in the "Peccioli: The Testing Site for the Robot DustCart" section, ubiquitous networking robotics in urban settings [6], and CyberCars [7]. However, evidence is not only limited to the field of research but also there are a few examples of commercial enterprises that have autonomous vehicles in their product catalogs, such as the French company [8].

The dimension of social pervasiveness deals with the potential impact of a given technology on the society. We argue that autonomous mobile robots may become a very pervasive technology in the near future. Because of their ability to move autonomously in the urban environment, robots can be designed to offer innovative and useful services to human beings. A few general examples of new applications that are still at research level are as follows: guiding people [6], support for the elderly and disabled [10], [11], and a mobile station for monitoring atmospheric pollution [1]. Another interesting field is the provision of solutions to well-known problems that affect urban areas [9], such as reducing road traffic by offering alternative services to people mobility [7], improving rubbish collection and transportation [1], or street cleaning [1]. In private settings, such as factories, there are many types of guide robots that are commercially available [12], [13].

However, a much more reliable indicator of social pervasiveness is given by the growing international market associated with service robotics. Drawing on the figures made available by the International Federation of Robotics in 2010 [14], till 2009, about 77,000 service robots for professional use were sold worldwide and the total value of professional service robots sold was about US$13 billion. In projections for the period 2010–2013, about 80,000 new service robots for professional use will be installed, and the estimated value of sales of service robots for professional use is estimated to be more than US$12 billion. About 30% of the sold units are used for defense applications, 25% for milking, 8% for cleaning, 8% for medical purpose, 7% for underwater robots, 6% each for demolition robots and mobile robot platforms for general use, 5% for logistic systems, and 4% for rescue robots. All these values indicate that robots will become a vital part of our daily life.

Peccioli: The Testing Site for the Robot DustCart

The testing of DustCart included the implementation of the DustBot system, actually, a slightly modified version of the system developed in the framework of the DustBot project [1]. The DustBot system has already been described in detail [15]. Unlike previous demonstrations of the DustBot system, which lasted only a day and were substantially structured, in Peccioli, for the first time, the system was tested in a real operative environment and with real users. The objectives of the test were to assess the performance of the system, identify its limits, technological as well as those related to the acceptance of the service and the robot by the end users, and to evaluate the economic sustainability of the whole system. The resulting data were necessary to evaluate the feasibility of the industrialization of the DustBot system. In this section, we describe the main elements that made the testing of the robot possible.

The DustCart Robots

The objects of the test were two robot prototypes of the DustCart, which is a mobile autonomous service robot, designed to carry out door-to-door, separate waste collection on demand. The robot consists of a mobile platform, originally the robotic mobility platform 200, a two-wheel robot commercialized by Segway, and customized to support a bin container for the transport and discharge of waste. To make the robot safer and increase its endurance, however, a new mobile base was developed from the scratch for the test in Peccioli. The new base consists of two actuated wheels and two supporting wheels, which overcome a few technical limitations intrinsic in the Segway platform. The robot was also equipped with two additional powerful batteries that allowed it to work continually for about ten hours. Thanks to the presence of special sensors and other components [16], the robot was able to navigate autonomously avoiding obstacles while moving. With regard to human–robot interfaces, the interaction with human beings took place by means of a touch-screen interface and consisted of simple operations: pressing a graphical button opened and closed the bin container and selecting the corresponding icon on the screen specified the type of waste to be disposed of. The whole interaction procedure was also accompanied by vocal messages. The robot was designed with the usability and acceptance criteria [17] in mind.

Figure 1. A view of Peccioli. (Photo courtesy of Giancarlo Teti.)

The Service Provided by DustCart

The service provided by DustCart during the test period in Peccioli was on-demand door-to-door waste collection. The robot was configured to collect three types of waste: undifferentiated, paper, and plastic. The service was in operation from Monday to Sunday, from 8:00 a.m. to 8:00 p.m. except on Tuesdays, when the service was in operation only from 3:00 p.m. to 8:00 p.m. owing to the local market present in the experimental area. To request the DustCart service, users had to call a toll-free number. The calls were managed automatically by the ambient intelligent (AmI) infrastructure (which is described later), which scheduled and allocated the robots. Once a robot was allocated the task, the AmI sent a short message service (SMS) to the user, informing him/her of the arrival time of the robot.

The Municipality of Peccioli

Peccioli is a historic village located in the countryside of Tuscany. The village, founded in the Middle Ages, was built on top of a hill and presents the classical topography of a medieval town, with old buildings and narrow, paved streets of various gradients (Figure 1). About 5,000 people live in Peccioli, with a large percentage of elderly people: about 25% of the inhabitants are more than 65 years old. The municipality of Peccioli has a strong penchant for using advanced technologies, with a view to provide its citizens with excellent public services. As a matter of fact, since 1995, the municipality of Peccioli has been collaborating on joint research projects with Scuola Superiore Sant'Anna (SSSA), and as a result, the citizens have had the opportunity to use and test very advanced research and experimentation facilities in the fields of aging, telemedicine, domotics, rehabilitation technologies, energy, environment, wellness, etc.

The Test Site

The test site covered three streets and a part of a square. This is the very heart of the town, which is also called *ciambellone* (which means doughnut) on account of its round shape (Figure 2). The total length of the path selected for the test was approximately 300 m. Within the test site, the streets are almost flat and paved, with shops, bars, restaurants, and other commercial activities on either side. It is worth noting that this is not a pedestrian area, as the roads can still be used for their conventional purpose. The busiest area of the town was selected as the test site to make the robots' presence strongly noticeable to people, so as to

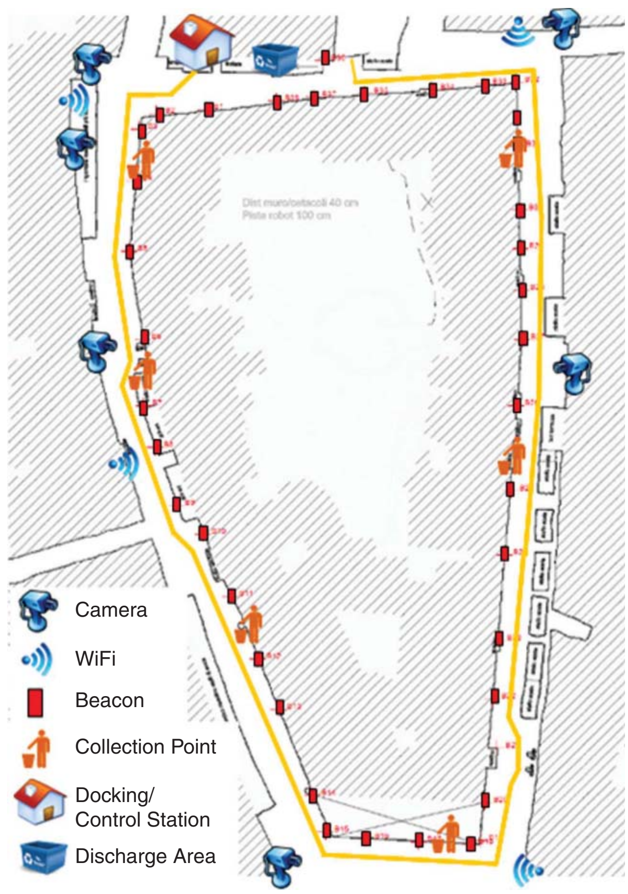


Figure 2. The experimental site with the location of the beacons, docking station, cameras, and access point of the wireless network. (Photo courtesy of Giancarlo Teti.)

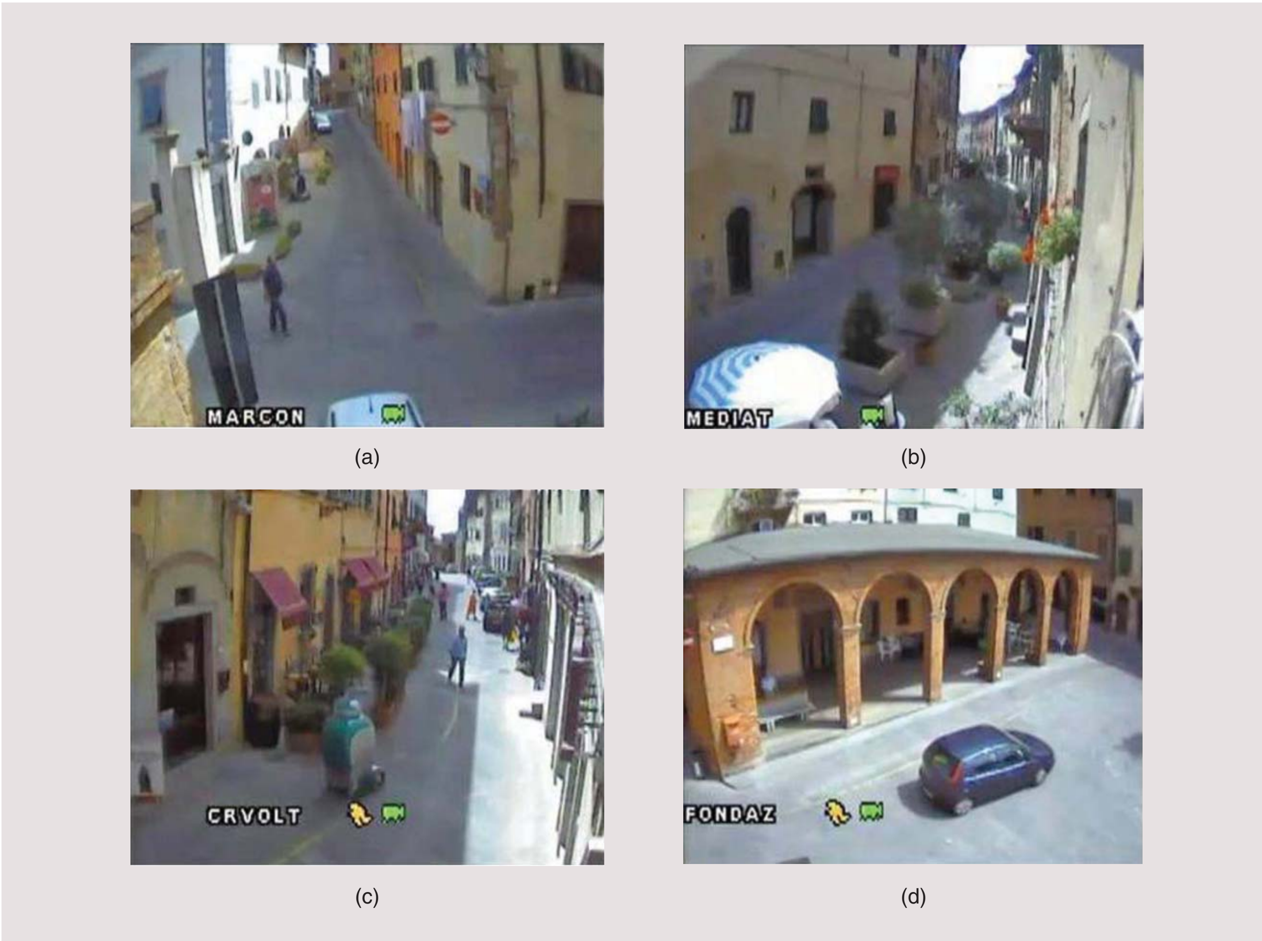


Figure 3. (a) and (b) Images from the four CCTV cameras used to supervise the robot. (c) The robot DustCart is in motion. (d) The images correspond to three streets and one square. (Photo courtesy of Giancarlo Teti.)

Figure 4. DustCart discharging a rubbish bag. (Photo courtesy of Giancarlo Teti.)

better evaluate the robot's social acceptance and give the word coexistence a concrete meaning.

Within the test site, the control and docking stations were also located. The control station is the place from where human operators supervise the functioning of the DustBot system. In the control station, human operators monitor the robots via remote real-time images that are sent from cameras positioned across the test site (Figure 3). The docking station is the place where the robots stay during the night and where they recharge their batteries and undergo maintenance operations (Figure 4). Near the docking station is the discharge area, which consists of a ramp with a dais, where the robot discharges the collected rubbish. The dais has three holes that correspond to the three typologies of waste collected by the robot (i.e., undifferentiated, paper, and plastic) (Figure 4).

A video surveillance system and a wireless network were installed in the testing site. Wireless coverage was assured by six access points located in the area and connected to the LAN of the control station. The surveillance system consists of six closed-circuit television (CCTV) cameras covering all the experimental areas and connected to a central recording system located in the control station.

The AmI

In the control station, a PC with a software named *AmI* for managing the system and supervising the activities of the robots was set up (Figure 5).

The users involved in the test were registered in the AmI software, and their telephone numbers were stored in the AmI database along with the robots and collection points. The software allowed us to associate the users to collection points. The AmI software communicates with the robots through a wireless network: when the AmI received a request by phone from a user, it scheduled the first free available robot and tasked the robot with serving the user. While moving, the position of the robot and its status is shown on the map. AmI also allows the operators to stop and resume the motion of the robot in the case of an emergency and/or to cancel the task.

Recruitment of Test Participants

The users for DustCart service were recruited from volunteers during a public assembly that was held before the test started and was open to the people of Peccioli. During the assembly, the system and how to use it was explained to the participants and, at the end, 34 users, which consisted of 18 families and 16 commercial activities, agreed to participate in the testing of DustCart. Participants agreed to use only the DustCart robot for the disposal of their rubbish bags for the whole duration of the test period. The majority of the users were 45% retired people and 35% workers. The average number of persons per family is 2.2, and the average age is about 52 years. Before the testing of DustCart, about 50% of the inhabitants of Peccioli separated domestic waste by using traditional bins located in specific areas of the town.

To determine the overall level of acceptance of innovations such as DustCart, it should also be necessary to include ethical, social, and legal issues.

Legal and Social Reflections

In addition to technological problems, the major difficulties that had to be overcome to make the testing of DustCart possible were those related to solving legal issues and avoiding social resistance. The following subsections focus on what was done, on the one hand, to solve legal problems related to the robot presence on public roads and, on the other, to improve the robot acceptability among people (Figure 6).

Legal Regulations

One of the first questions faced by the organizers of the testing was, "Is it possible to use robots on public roads?" From previous research and studies, we know that there

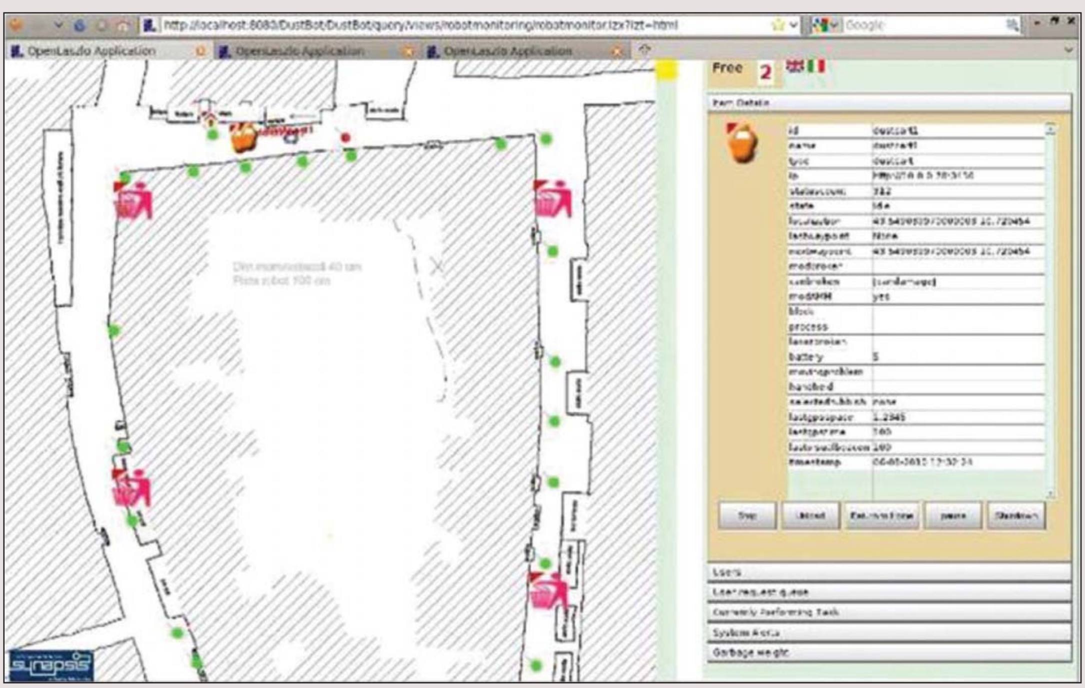


Figure 5. The graphic user interface was developed by Dedalus Srl. (Photo courtesy of Giancarlo Teti.)

Figure 6. A user giving rubbish to the robot. (Photo courtesy of Foto Silvi.)

exists a legal gap with regard to the juridical status of service robots operating on public roads [18], [19], at the European and presumably international levels. In Europe, according to Article 8 of the "Vienna Convention on Road Traffic," each moving vehicle, including animals, shall have a driver [20]. In other words, for the road traffic convention, including the Italian highway code, an autonomous vehicle is a contradiction in terms. In what follows, we will discuss the solutions that were adopted and as a result allowed DustCart to operate on the public roads of Peccioli. The municipality in collaboration with the local municipal police took the following measures to ensure safety, avoid traffic congestion, and allow the robot to accomplish its task.

- *Road signs*: Three new road signs were specifically designed for the testing of DustCart.
  - A) A general warning sign, highlighting the presence of robots operating on the streets. On the sign, there is the following text: "Attention. Area subject to robotic testing."
  - B) A more specific sign, informing the public of the presence of a yellow lane given over to the robot (Figure 7), in which the following text appears: "Attention. Area subject to robotic testing. Yellow lane used by robots."
  - C) Finally, a more specific sign used to warn road users of the possibility of robots crossing the road, which displays the following text: "Attention. Robot crossing. Yellow lane used by robots."
- *The robot lane*: This lane is a special yellow strip, drawn on the left-hand side of the roads (Figure 8), involved in the test. It was decided that the robots would travel inside the lane, in the direction as that of the cars. The robot lane was meant to separate the robot's activities from the traffic. It was also meant to prevent cars as well as bikes and bicycles from parking in the robot's path. To reduce traffic congestion on the narrow streets of Peccioli, three stops were devised on each road to give way to cars.
- *Robot insurance policies*: As the owner of the robots, SSSA was given the task of providing insurance cover for the robot for the duration of the test period. Our first move was to ask our broker whether we would be able to use our insurance policy to cover DustCart against any liability that resulted from the robot's labor (i.e., death or injury to people or animals). From past experience with public demos, we knew that the SSSA insurance policy covered all research activities, including demos, carried out with our prototypes by the institution personnel, anywhere in the world. However, this time, the broker told us that because of the peculiar nature of the event, the insurance company requested a specific appendix for testing DustCart and the payment of an additional insurance premium. The reason for these requests was the difficulty in placing our robot DustCart within a specific typology transport identified by the Italian highway code,

Figure 7. Example of a new road sign (C type), which was specifically designed for the testing of DustCart in Peccioli, Italy. (Photo courtesy of Foto Silvi.)

and this was due to the lack of a human driver. Hence, the robots cannot be subject to the obligatory insurance that covers damage when operating on the road. Eventually, we managed to insure the robot for the test by paying an additional insurance premium of about €800. However, the policy did not cover damages that the robots may sustain. According to our broker, this kind of risk seems to be the most difficult to cover for a company in the insurance market in economically and normative advantageous conditions.

- *Privacy*: Robots, such as DustCart, are endowed with the ability to perceive, store, and use sensitive data related not only to the environment but also to human beings. Therefore, as pointed out by [9], there exists a privacy problem for robots operating in public areas. In Peccioli, the privacy problem was solved by placing in the test site signs warning of the presence of cameras, which were used for security reasons and in accordance with Article 13 of Italian "Personal Data Protection Code" (30 June 2003).

Figure 8. The robot lane (right-hand side of the picture). (Photo courtesy of Giancarlo Teti.)

Social Acceptance

With regard to the presence of an innovation such as DustCart robot in an urban community, our experience shows that social acceptance in this case depended on the following:

1) citizens' perspective
2) the adoption of a specific innovation process by the local municipality, which was fully supported by our team.

The innovation process—from research to the diffusion and adaptation phases—should take into account the users' perspectives, in a twofold way. On the one hand, users will adopt (or not) an innovation, awarding its success. On the other hand, they will undergo an innovation, sometimes over their original needs and expectations. Users' acceptance of innovations such as DustCart, when proposed to an urban community, is also related to two factors: the public's trust in the local political government and the ability to enable users (citizens) to acquire a balanced view of the new methods required to carry out something they had done differently before. According to the technology acceptance model (TAM) [21], the user's acceptance of any technology depends on two factors that have a significant impact on a user's attitude toward using the technology: perceived usefulness and perceived ease of use.

These considerations were taken into account during the two public meetings that were organized by the municipality of Peccioli to involve the citizens in the testing of DustCart. In addition, during the public meetings, the five stages in the decision process identified by [21] were taken into account. As to knowledge and persuasion, all Peccioli citizens were duly informed of the DustCart test. For this purpose, the first public meeting was organized, whose objective was to provide inhabitants of Peccioli with a wide range of information concerning the project and the possible impact of this innovation on Peccioli's future.

For the same purpose, the organizers carried out a press campaign that included journals and magazines. As a result, some citizens got interested in the test of the robot and actively sought further details about the innovation.

The relationship between government and citizens' decisions has been found to be very important in the adoption of the robot and its service.

For what concerns decision, citizens evaluated the advantages/disadvantages of using the innovation DustCart and decided whether to possibly adopt or definitely reject the innovation. In the testing area, there were 110 users (families and commercial premises): some of them asked to be involved in the test (early adopters [22]), and the first 24, according to the DustBot potentialities, were accepted. An interesting point was the fact that the link with tradition played an important role in accepting the innovation DustCart: people who showed some initial doubts about the robot were definitely convinced as soon as someone remembered them the ancient dustman, named *Oscar*, who used to collect garbage in a truck when called by Peccioli citizens.

Figure 9. A curious and interested group of people.

At the implementation stage, usually the individual employs the innovation to a varying degree that depends on the situation and is determined by the usefulness of the innovation and may possibly search for further information about it. In our case, the local municipality organized a second public meeting; this time, the meeting included training on how to use the DustCart robot and request its service and information on its usefulness and ease of use.

Finally, as to confirmation, the citizens finalized their decision to use DustCart for the entire period of the test, using it to its full potential.

By investigating the impact of political factors upon urban policy outputs, the relationship between government and citizens' decisions has been found to be very important in the adoption of the robot and its service. In this process, the organizational environment and the local government were significant variables and directly influenced the testing of the innovation DustCart, where many other predictors of the use of this innovation would have been expected to be more important. These factors cannot explain or predict the success or failure of the future use of DustCart, but they certainly played a fundamental role in this test.

Conclusions

During the test period, the robot operated for 47 days (from 15 June 2010 to 7 August 2010), for a total of 454 h (with reduced service on Tuesdays), carrying out 382 services and traveling a total of 114.6 km. A total of 560.3 kg of rubbish was collected (paper: 226.5 kg, 40.42%; plastic: 89.7 kg, 16.01%; and undifferentiated garbage: 244.1 kg, 43.57%). During the testing of DustCart, no accidents occurred, and the robot proved to be reliable and safe.

During the testing of DustCart, no accidents occurred, and the robot proved to be reliable and safe.

In this article, we have attempted to highlight that there also exist nontechnological challenges for deploying service robots in urban areas. In particular, we have pointed out the implications related to legal regulations and social acceptance.

As far as the legal issue is concerned, as we have pointed out, the legal classification of autonomous robots at the level of road traffic code still remains a problem. The solutions that were adopted in Peccioli are valid only for a temporary test. With regard to social acceptance, we believe that DustCart was accepted by the inhabitants of Peccioli, and this was on account of the fact that people associated its presence with the accomplishment of an important service: separate waste collection (Figure 9). Citizens also appreciated the fact that DustCart offered an on-demand service; in other words, it fitted the needs of each person and provided a door-to-door service, which meant that people did not have to move away from their homes or shops to dispose of rubbish. This last feature was quite important, especially, for elderly people and for those working in shops.

However, although not a demonstration, but a real deployment, the testing of DustCart was limited in time and took place under partially controlled conditions. In other words, there was neither sufficient time nor the appropriate conditions to find out other potential ethical, legal, and social implications. For instance, it was not possible to find out the existence of abuses or improper behaviors toward the robot, such as vandalism. It has been discussed earlier [23] that vandalism can affect the safety as well as quality of service provided by robots operating in urban environments. A very simple example of robot vandalism is a group of people surrounding the robot and blocking its way. In addition, it would have been interesting to evaluate dustmen's acceptance level in case the robot was used permanently or that of road users. As discussed earlier, one of the main problems encountered during the testing of the robot was traffic congestion. Despite the presence of a robot lane, the streets of Peccioli are too small for both robots and cars. In conclusion, among the lessons learned from Peccioli that are useful for paving the way for robot deployment in urban settings is that acceptance is no more a matter of the users willingness to use a product [24]. On the contrary, because the robot coexists in a public environment, to determine the overall level of acceptance of innovations, such as DustCart, it should also be necessary to include ethical, social, and legal issues.

Acknowledgments

The DustCart robot was originally developed in the framework of the European Union-funded project DustBot. The authors are grateful to all the people involved in the organization and realization of the DustCart test, especially Belvedere S.p.A. and its president Renzo Macelloni, the mayor of Peccioli Silvano Crecchi, Peccioli municipal police, Scuola Superiore Sant'Anna, the operators in the control station Silvia Fornai and Marco Toncelli, and RoboTech Srl and Dedalus, the tow spin-off companies of SSSA that optimized and adapted the hardware and software components of the DustBot system for the test. Finally, a special thanks to all the citizens of Peccioli who supported and participated in the project.

References

[1] A. Proyas, *I Robot*. Twentieth Century Fox Film Corporation, U.S./Germany, 2004.

[2] G. Tamburrini, "Brain to computer communication: Ethical perspectives on interaction models," *Neuroethics*, vol. 2, no. 3, pp. 137–149, 2009.

[3] (2010, July 27). Available: http://ethicbots.na.infn.it/kickoff.htm

[4] (2010, July 27). Available: http://www.darpa.mil/grandchallenge/index.asp

[5] Information Society Technology. (2010, July 6). DustBot project—Networked and cooperating robots for urban hygiene, collaborative research project, FP6 [Online]. Available: http://www.dustbot.org

[6] Information Society Technology. (2010, July 6). URUS project Ubiquitous networking robotics in urban Settings, collaborative research project, FP6 [Online]. Available: http://www.urus.upc.es/

[7] Information Society Technology. (2010, July 6). CyberCars project Cybernetic cars for a new transportation system in the cities, collaborative research project, FP6 [Online]. Available: http://www.cybercars.org/

[8] Technopole d'Izarbel, France. (2010, July 6). RoboSoft advanced robotics solutions for transport, cleanliness, security and health [Online]. Available: http://www.robosoft.fr.org/

[9] A. Sanfeliu, A. Punsola, Y. Yoshimura, M. R. Llácer, and M. D. Gramunt, "Legal challenges for networking robots deployment in European urban areas: The privacy issue," in *Proc. Workshop Network Robots Systems, IEEE Int. Conf. Robotics and Automation (ICRA 2009)*, Kobe, Japan, May 12, 2009, pp. 1–14.

[10] E. Prassler, J. Scholz, and P. Fiorini, "A robotics wheelchair for crowded public environment," *IEEE Robot. Automat. Mag.*, vol. 7, no. 1, pp. 38–45, Mar. 2001.

[11] C. Gao, M. Sands, and J. R. Spletzer, "Towards autonomous wheelchair systems in urban environments," in *Field and Service Robotics: Results of the 7th International Conference, Springer Tracts in Advanced Robotics*, A. Howard, K. Iagnemma, and A. Kelly, Eds. Berlin, Heidelberg: Springer-Verlag, 2010.

[12] Mobile Robots Inc. Autonomous mobile robot cores, bases, and accessories [Online]. Available: http://www.mobilerobots.com/RobotApplications/FollowMe_Workstations.aspx

[13] Mobile Robots Inc. Autonomous mobile robot cores, bases, and accessories [Online]. Available: http://www.mobilerobots.com/RobotApplications/Intelligent_AGV_Delivery_Robots.aspx

[14] IFR International Federation of Robotics. (2010, Dec. 20). [Online]. Available: http://www.ifr.org/news/ifr-press-release/service-robots-getting-successfully-established-192/

[15] B. Mazzolai, V. Mattoli, C. Laschi, P. Salvini, G. Ferri, G. Ciaravella, and P. Dario, "Networked and cooperating robots for urban hygiene: The EU funded DustBot project," in *Proc. 5th Int. Conf. Ubiquitous Robots and Ambient Intelligence (URAI)*, Seoul, South Korea, Nov. 20–22, 2008.

[16] G. Ferri, A. Mondini, A. Manzi, B. Mazzolai, C. Laschi, V. Mattoli, M. Reggente, T. Stoyanov, A. Lilienthal, M. Lettere, and P. Dario, "DustCart, a mobile robot for urban environments: experiments of pollution monitoring and mapping during autonomous navigation in urban scenarios," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA), Workshop Networked and Mobile Robot Olfaction in Natural, Dynamic Environments*, Anchorage, AK, May 7, 2010, pp. 1–15.

[17] P. Salvini, A. Armato, I. Mannari, B. Mazzolai, C. Laschi, and P. Dario, "A methodology for designing acceptability in relation to a service robot physical appearance," in *Proc. 8th Int. Workshop Human-Friendly Welfare Robotic Systems, KAIST*, Daejeon, Korea, Oct. 21–23, 2007, pp. 1–8.

[18] Y.-H. Weng, C.-H. Chen, and C.-T. Sun. (2007, June 4–8). The legal crisis of next generation robots: On safety intelligence. presented at ACM 11th Int. Conf. Artificial Intelligence and Law (ICAIL'07) [Online]. Available: http://works.bepress.com/weng_yueh_hsuan/2

[19] P. Salvini, G. Teti, E. Spadoni, E. Frediani, S. Boccalatte, L. Nocco, B. Mazzolai, C. Laschi, G. Comandé, E. Rossi, P. Carrozza, and P. Dario, "On robots and legal regulations: An investigation on how the law is impacting the deployment of robots in urban environments," *Adv. Robot.*, vol. 24, no. 13, pp. 1901–1917, 2010.

[20] (2010, July 20). Article 8 of the Vienna Convention on Road Traffic, 1968. Available: http://www.unece.org/trans/conventn/legalinst_08_RTRSS_RT1968.html

[21] F. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quart.*, vol. 13, no. 3, pp. 319–340, Sept. 1989.

[22] E. M. Rogers, *Diffusion of Innovation*, 5th ed. New York, NY: Free Press, 1995.

[23] P. Salvini, G. Ciaravella, W. Yu, G. Ferri, A. Manzi, C. Laschi, B. Mazzolai, R. Oh, and P. Dario, "How safe are robots in urban environments? Bullying a service robot," in *Proc. 19th IEEE Int. Symp. Robot and Human Interactive Communication, Ro-Man 2010*, Viareggio, Italy, Sept. 12–15, 2010, pp. 1–7.

[24] A. Dillon, "User acceptance of information technology," in *Encyclopaedia of Human Factors and Ergonomics*, W. Karwowski, Ed. London: Taylor & Francis, 2001.

Pericle Salvini, Polo Sant' Anna Valdera, Viale Rinaldo Piaggio, 34, 56025, Pontedera (PI), Italy. E-mail: p.salvini@sssup.it.

Giancarlo Teti, RoboTech Srl, Scuola Superiore Sant'Anna, Pisa, Italy.

Enza Spadoni, HealthCare and the Biomedical Field at BioRobotics Institute of the Scuola Superiore Sant'Anna, Pisa, Italy.

Cecilia Laschi, Scuola Superiore Sant'Anna (SSSA), Pisa, Italy.

Barbara Mazzolai, Center for Micro-BioRobotics (IIT@SSSA), Istituto Italiano di Tecnologia (IIT), Genoa, Italy.

Paolo Dario, BioRobotics Institute, Scuola Superiore Sant'Anna (SSSA), Pisa, Italy.

Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

Remote-Control Crimes

By Peter M. Asaro

Roboethics and Legal Jurisdictions of Tele-Agency

This article considers some of the potential legal implications of teleoperated robotic systems for enabling action at a distance or tele-agency. In particular, it considers issues that may confront law enforcement as well as issues of legal jurisdiction when tele-agency extends across the traditional physical boundaries of legal jurisdiction.

The legal approach is one of the approaches for the issues faced by roboethics. A consideration of ethics in robotics using the tools offered by the practice of law has been made elsewhere [1]–[3] and has focused on product liability and robots as legal agents. The difficulties of applying the law to some of the possible activities involving new robotic capabilities that may arise in the near future are considered.

One new capability, in particular, is that robotic systems also pose one of the greatest threats of social disruption. This new capability has, however, been largely overlooked by the rather small literature on roboethics, namely, the ability of robotic systems to support action at a distance, known as *tele-agency*. This capability has serious implications for both law enforcement and legal jurisdiction, though tele-agency has received more attention in the art world (e.g., [4], [5]) to date than it has in discussions of robot ethics and law. This essay seeks to correct this by considering some of the legal issues that might arise as teleoperated robots proliferate and spread into consumer markets, international trade, and hacker communities.

Digital Object Identifier 10.1109/MRA.2010.940154
Date of publication: 14 April 2011

Telecrimes and Law Enforcement Issues

Simply put, teleoperated and remotely controlled systems allow the legally responsible actor(s) in control of the system to be spatially (and, in the case of preprogrammed systems, temporally) distant from the effects of their actions, without requiring the support of human accomplices. This has several serious consequences for law enforcement because of the perpetrator's reduced bodily risk, the risk of being arrested in the conduct of a crime, and the difficulties

1070-9932/11/$26.00©2011 IEEE

involved in correctly identifying the responsible individuals. These are clearly matters of law enforcement—identifying, apprehending, and convicting the perpetrator of a crime—and do not affect whether or not the perpetrators are guilty of a crime. However, the ability to more easily commit crimes while reducing the risks of facing punishment is certainly a threat to justice and the public good.

While, in theory, there is little legal difference between robbing a bank at gunpoint and using an armed robot to rob a bank, there are significant practical differences. Most obviously, there is a significant difference in the bodily risks assumed by an armed robber and someone controlling an armed robot remotely, which has implications for the use of armed police and security guards as a deterrent to such crimes. Viewed another way, this could have a significant impact on the ability of police and security forces to intervene and stop such crime, especially if they are unable to physically subdue the robotic system. On the other hand, law enforcement could use force against the robot without the same restraint that would be called for if a human body were at risk (even the body of a suspected criminal).

The use of such a remote-robbing robot also requires the police to do additional work to correctly identify and locate the perpetrator of the robbery. If the current state of the art of cybercrimes is a good indication, it will likely be quite difficult to track down the perpetrator of such a crime when their control path has been routed through a series of networks and servers intentionally designed to obscure the identity and location of the criminal. Teleoperation implies the real-time control of a robot from a physical distance. We can, however, also consider a preprogrammed robot as a kind of teleoperation in which the programmer/controller is temporally removed from the actions of the robot, though is still the responsible agent. Again, there are precedents for treating programs as a form of criminal behavior, as is done in the creation and use of illegal viruses and botnets.

Robotic technologies might not prove to be a source for a massive remote-control crime wave, however. Some reasons for this are that, first, in material-property theft, the stolen property must be recovered at some point, and thus the police could track and follow the robot until the perpetrators attempt to retrieve the property. Telerobotic theft is quite unlike cybertheft in this regard, as stolen information can be quickly and easily transferred through the network, while stolen material objects cannot. Second, initially such robotic technologies will be expensive and thus would be unlikely to be used in petty crimes where the value of the stolen goods is less than the cost of losing the robot. Third, like other cybercrimes, these tele-agency crimes might leave data trails that could be used to identify perpetrators, and videofeeds and control commands might actually be recorded by authorities in ways that could be used in courts to prove the guilt of perpetrators. While it is already possible to commit complex cybercrimes, robotic technologies will extend the range of these crimes into the embodied material world, including bodily violations and violent crimes such as assault, rape, and murder.

Tele-Agency Across Jurisdictions

Complex legal questions may also arise when the perpetrator controlling the system and the robot being controlled are in different legal jurisdictions. Certain interjurisdictional or multijurisdictional actions are already handled by the law in various ways. Examples include using other human agents to conduct a crime, such as in conspiracy or being an accomplice to a crime, though these tend to carry lesser penalties than the actual commission of the crime. In the United States, a crime (e.g., a fraud) that involves actions in two or more different states within the country can result in the matter being settled by the federal court system rather than in the state courts. Prosecutors may also seek convictions for crimes in each state jurisdiction separately, depending on the case and cooperation between state and federal prosecutors. In these cases, there are often similar sets of laws that apply in each jurisdiction, and sometimes one set overrides, such as federal law having precedence over state, provincial, or local laws. More controversial are cases in which there are subtle differences in the definitions of what constitutes the crime in question or when different penalties may apply, depending on the court and jurisdiction in which the trial is held. For example, a first-degree murder in some U.S. states carries a death penalty, but not in others, and so it can matter a great deal where the crime is committed and tried. Indeed, the rules of extradition in some jurisdictions, as in many countries in Europe, are such that they will not allow the extradition of a suspect for trial in another country in which they might face the death penalty. More generally, extradition requires an international treaty agreement, which usually stipulates that the charges be serious enough to warrant returning an individual, such that many petty crimes committed using robots might not warrant extradition.

These systems present a new capability for committing violent crimes at great distances that did not exist before.

The more problematic cases involve activities that are legal in one place but illegal in another. A good example of this is gambling. In most states, gambling is illegal, or at least tightly regulated by the state. With the advent of the Internet, however, it became possible to engage in gambling activities online. The legal question then arose as to where the gambling is taking place. If the gambler and the computer server running the gambling program are both in a jurisdiction where gambling is legal and the activity is properly licensed, then the activity is legal. But is it still legal when the player is in a jurisdiction where it is not legal to gamble but the server is? What if two gamblers are betting in the same poker game, but one is in a jurisdiction where it is legal and the other is not? Is one engaged in an illegal activity but the other not, even though they are playing the same game? What if the players are in a legal gambling jurisdiction but the server is not, or the network passes gambling-related data traffic through computers in a jurisdiction where it is illegal? While we could propose simple and consistent legal interpretations, e.g., that what matters is where the gamblers are, it does not mean that the courts are necessarily free to apply them. They must also weigh issues of public interest, legal precedent, and often, the decisions of other courts, or even the treaties and legal bodies that constitute international law.

A promising legal concept that might serve to prevent the use of robotic technologies to exploit local differences in legal standards is the universality principle or universal jurisdiction.

The issue of tele-agency and gambling has, in fact, been addressed explicitly, not by a court exactly, but by the dispute-resolving mechanisms of the World Trade Organization (WTO). The WTO is a multiparty international treaty organization whose rules and decisions are binding upon member nations. In 2003, the small island nation of Antigua petitioned the WTO against the United States for their enforcement of antigambling laws on gamblers within the United States who logged in to computer servers in Antigua to engage in gambling [6], [7]. Antigua argued that the actions of the United States in enforcing those laws hurt their ability to engage freely in trade with a market of gamblers in the United States. As the servers were in Antigua, they argued that the gambling was in Antigua, and the United States was engaged in protectionism by denying those players the opportunity to engage in free trade with legal businesses in Antigua. The United States argued that permitting players in the United States to gamble online undermined their ability to use the law to enforce a public moral interest and to maintain social control within its borders. In 2005, the WTO ruled in favor of Antigua.

In accepting this argument, the WTO effectively legalized online gambling in all WTO member nations, provided that gamblers used computer servers located in a jurisdiction that is a member of the WTO and in which gambling was legal. The effective modifier here is significant, because the WTO is not really a court, and a WTO member nation could still choose to enforce their antigambling laws, though they would be subject to WTO penalties and fines for protectionism, or they could withdraw from the WTO altogether. It is also important to note that the basis of this decision is not simply that online gambling is legal, because the servers are in Antigua where gambling is legal. This is unlike other precedents in international law for two important reasons. First, because the WTO does not have legal authority beyond its member nations, and thus, its legal decisions do not carry the weight of precedent that, e.g., the decisions of the International Criminal Court would. Second, because the WTO only has authority over international trade, future use of this decision as a precedent would only be applicable in other cases involving free trade and protectionism among member nations.

That said, we can envision a variety of possible scenarios in which an online activity involved the use of teleoperated robotics and was a matter of free trade. That is, where the activity involved would be illegal if it were engaged in locally, but a commercial industry might exist in which people were willing to pay for the opportunity to circumvent local laws through remote teleoperation and thus the WTO decision would apply as a precedent. For example, in 2004, a Texas entrepreneur launched a Web site (www.live-shot.com) that, for a fee, allowed users to log in, aim, and fire a real gun at real targets. His ultimate plan was to provide live animals for a teleoperated hunting business, claiming that this would serve a market of physically impaired hunting enthusiasts who could not go out into the woods themselves [8], [9]. The business and Web site are now defunct, because 11 states including Texas passed laws making online hunting illegal by requiring the hunter to be physically present when hunting.

Interestingly, the Texas law prohibits anyone hunting with a robot within the boundaries of the state but would not necessarily apply to hunters in Texas going online to, e.g., hunt big game in Africa with a robot. If the laws were written so as to prohibit the act of online hunting itself then the online gambling precedent would apply. If the hunting range was set up in Antigua, for instance, and represented a legal and profitable business interest in Antigua, laws prohibiting online hunting in Texas would be unenforceable due to the WTO ruling.

It is worth noting that the morally abhorrent nature of the activity in the gambling case was not sufficient to justify enforcing the local laws over promoting free international trade [6], though perhaps some activities could reach a level of abhorrence that this would no longer be true. The animal hunting case would not seem to rise to such a level. But we could imagine jurisdictions that either lacked certain legal prohibitions or decided to permit certain activities to generate trade revenue by attracting customers wanting to engage in teleoperated activities, precisely because they are illegal or prohibited in the locales where their online customers reside. Because of this, the commercially successful activities are likely to descend toward the questionable and prurient end of the moral spectrum, including sexual acts, violent acts toward animals and humans, or human degradation and torture. This raises a disturbing set of questions: What if there was a jurisdiction willing to sell the opportunity to execute people who have been sentenced to death? or which allowed humans to consent to risking their own lives in mortal combat with teleoperated robots? And, as human slavery remains marginally legal in a handful of countries, would the consent of slave owners be sufficient to decriminalize physical violence up to death against a slave by a robot operator in another country in which such slavery and violence is criminal? In a globalized economy that has already seen banks and multinational corporations establish offices or incorporate in jurisdictions that offer them the greatest protection from taxes or other legal liabilities or restrictions, we should not be surprised when certain locales seek to enrich themselves by becoming safe havens supporting the circumvention of other established legal jurisdictions. Should robotic crimes fall through similar jurisdictional cracks as online gambling and offshore tax havens, we might well see the emergence of some sort of robot safe-havens.

Beyond the economic and trade aspects, there are critical issues of interjurisdictional enforcement as well. Even if it were acknowledged that an illegal act was committed in jurisdiction A, using a robot being controlled by a clearly identified person in jurisdiction B, it is not clear that courts in jurisdiction B would necessarily be able to prosecute the offender. Jurisdiction A would first need the person to be arrested and extradited from jurisdiction B to prosecute them, but not all countries have extradition treaties with each other, and, even where there are treaties, not all crimes warrant arrest or extradition.

A promising legal concept that might serve to prevent the use of robotic technologies to exploit local differences in legal standards, or circumvent prosecution due to jurisdictional gaps, is the universality principle or *universal jurisdiction* [10]. It was famously used by Spanish courts to arrest and charge Augusto Pinochet for crimes he committed as the dictator of Chile, though he never stood trial. However, the justification for applying the principle of universal jurisdiction is that the crimes committed are so heinous that they are crimes against all of humanity, and thus, all courts have the authority to prosecute suspected offenders. It is doubtful that most specific cases involving robots would rise to the level of crimes against humanity. It is also doubtful that Spanish courts (or the Belgian courts that also assert universal jurisdiction) would have an interest in prosecuting tele-agency crimes around the world or have the resources to do so. We might instead imagine that certain specialized courts could be constituted and supported by international treaty organizations, such as the United Nations, or by perhaps expanding the scope of crimes considered by the International Criminal Court. But such a development would likely arise in the face of public outcry in multiple countries at the inability of existing legal structures to reign in a growing number of such crimes.

Conclusions

In summary, the development and use of teleoperated robotic systems will continue to present new difficulties for the enforcement of local and international laws. These systems present a new capability for committing violent crimes at great distances that did not exist before. Moreover, the ability of tele-agency to separate actors from their actions will further enable the exploitation of inconsistencies between the legal standards of different jurisdictions. These legal issues are likely to be exacerbated by recent developments in international trade and globalization. There are some counterweights to these rather bleak possibilities however. First, robots only provide a margin of anonymity to their controller and not complete anonymity. Second, there are fundamental asymmetries in tele-agency, such that information can be transmitted in both directions, but material entities and properties are stuck on the effector end of the robotic system. Finally, the jurisdiction issues could be addressed by international courts or universal jurisdiction, but the establishment of such courts is unlikely, and most cases of telerobotic crimes will fail to rise to the current high standards set for universal jurisdiction.

References

[1] P. M. Asaro, "What should we want from a robot ethic?," *Int. Rev. Inform. Ethics*, vol. 6, no. 12, pp. 9–16. Reprinted in R. Capurro and M. Nagenborg, Eds., *Ethics and Robotics*. Amsterdam, The Netherlands: IOS Press, 2009, pp. 1–10.

[2] P. Asaro. (2007, Apr. 14). Robots and responsibility from a legal perspective. *Proc. IEEE Conf. Robotics and Automation Workshop on Roboethics*, Rome, Italy [Online]. Available: http://peterasaro.org/writing/ASARO%20Legal%20Perspective.pdf

[3] P. M. Asaro, "A body to kick, but still no soul to damn: Legal perspectives on robotics," in P. Lin, K. Abney, and G. Bekey, Eds., *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge, MA: MIT Press, to be published.

[4] K. Goldberg, Ed. *The Robot in the Garden: Telerobotics and Telepistemology in the Age of the Internet*. Cambridge, MA: MIT Press, 2000.

[5] E. A. Shanken, "Tele-agency: Telematics, telerobotics and the art of meaning," *Art J*, vol. 59, no. 2, pp. 64–77, Summer 2000.

[6] T. Madger, "Gambling, the WTO, and public morals: A short review of Antigua versus the United States," *Televis. New Media*, vol. 7, no. 1, pp. 52–67, Feb. 2006.

[7] Antigua On-line Gambling Association. [Online]. http://www.antiguawto.com/WTODispPg.html

[8] J. Cummings. (2004, Nov. 23). Click–bang! Online hunting faces hurdles: Inventive hunter hopes to give new meaning to 'point and shoot.' *NBC News* [Online]. Available: http://www.msnbc.msn.com/id/6568977/

[9] M. Matthew. (2005, Sept. 28). State lawmakers bag online hunting. Stateline.org [Online]. Available: http://www.stateline.org/live/Viewpage.action?siteNodeId=Id56691

[10] K. C. Randall, "Universal jurisdiction under international law," *Texas Law Rev*, vol. 66, no. 4, pp. 785–842, 1987.

Peter M. Asaro, Department of Media Studies and Film at the New School University, New York, NY, 10011, USA. E-mail: asarop@newschool.edu.

IEEE Robotics & Automation Magazine | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

Ethics in Advanced Robotics

ELS Issues in Advanced Robotics

By Fiorella Operto

> The meanings of ethical concepts and rules, in a given situation, should be clear and unambiguous. If they are not, one must undertake to clarify their meanings to the extent possible (..) New ethical judgments and cases should be assimilated, where possible, into the existing body of cases, rules, laws, policies, and practices.
>
> Terrell Ward Bynum on Norbert Wiener

Since 2003, the ethical, legal, and societal issues (ELS) in advanced robotics have attracted the increasing and lively interest of academic and professional circles. A similar, although more occasional, debate has also spread to the general public, stimulated either by the novel statements of researchers about recent advancement in robotics or by new and sensitive robotics applications. This increase of contributions and interest occurred hand in hand with the rapid development of research and applications in the service (personal) robot sector, marking the end of the robot segregation era [1].

Many are the current social motivations for a high demand for personal robots. On the economic side, the transitional conclusion of a bull cycle for

Digital Object Identifier 10.1109/MRA.2010.940155
Date of publication: 14 April 2011

1070-9932/11/$26.00©2011 IEEE

industrial robotics; on the social side, the new demand for dependable and safety automatic autonomous machines to be employed in human assistance. Although the so-called social robots have not yet been sold as standardized market products for consumers, we can foresee that the research and application will point in this direction, driven both by the market demand and the challenges and richness of technological and scientific issues posed to the roboticist researchers (see the concept of social caretaking [2]).

Some major segments of our lives already depend on sophisticated machines. However, in some critical instances (where robots are entrusted with human lives, such as in medicine, human care, or war theaters), we are approaching an ethical and regulatory gap, primarily because of the lack of new criteria to ascribe responsibility to software agents and learning machines.

In fact, the recent thrust in research and applications programs aiming toward developing robots that are able to cooperate with humans—that are, software agents and robots entrusted with high learning and decision-making functions—presents many theoretical and practical cases in which moral and legal aspects of the responsibility-attribution problem for learning robots may soon become an urgent issue [3]. The responsibility-ascription problem is a central issue in human-robot interaction (HRI).

These and other epistemic limitations and gaps in our current moral codes and regulatory guidelines, concerning software agents, learning machines, and robots, required, in the view of roboticist Gianmarco Veruggio—who coined the word Roboethics and started the debate on this issue—an effort of analysis and original research in this new field of ethics in science and technology.

Roboethics is the applied ethics developed for 1) the identification and analysis of ethical issues that arise from current and upcoming robotics applications and 2) the possible definition of some general guidelines on the issues mentioned earlier. In a broader meaning of the word, roboethics not only applies to the so-called negative rights (that is, the prohibition of some actions) but also the identification of some trends of robotics research and development that enhance what are viewed as human positive rights (the encouragement to some actions): robotics research and applications that promote and enhance human rights for well-being, education, right of medical care, food, housing, and employment. Actually, robotics and bionics systems may potentially make a significant contribution to the solution of many open human issues. Assistive robots could promote the right of people to live a life of independence and social participation; robotic systems in medicine could foster fundamental human rights by improving the quality of medical operations and protecting the patient's physical integrity; biorobotics have the potential to provide effective therapeutic means to restore lost motor functionalities; and robots for the environment could be extremely useful in the cleaning of polluted areas.

Following the first International Symposium on Roboethics [5], one of the first steps in roboethics was to select a methodology for the identification and analysis of technoethical issues in robotics. To our knowledge, three European projects (which also availed themselves of international experts as well as background studies in information and computer ethics and bioethics) represent the most structured efforts in this field. These are as follows: the Euron Roboethics Atelier, 2006 [6]; Ethicbots 2006–2008 [7]; and Coordination Action for Robotics in Europe (CARE) 2008 [8]. Here, the analyzers could 1) identify the most universally shared and enforced declarations, conventions, and agreements about human rights; 2) achieve a shared knowledge about the ethical notions involved in the potential violation and promotion of human right in relation to human interaction with robotic, bionic, and artificial intelligence (AI) systems; and 3) apply their contextual understanding of these ethical notions to some technoethical cases. This article has adopted the same methodology.

The further step was to identify the human rights to be protected and promoted by roboethics, listed and discussed in the Universal Declaration of Human Rights (1948) and the Treaty of Lisbon (2000). These are as follows:

- respect and protection of human dignity and privacy
- right for physical and integrity of the person
- right for liberty and security
- right to protection of personal data
- right for the elderly to lead a life of dignity and independence and to participate in the social and cultural life
- right to integration of persons with disabilities
- fair access to technological resources and social and cultural discrimination (per ages, gender, and census).

A further phase resulted in applying roboethics analysis to the following cases:

- *Human dignity and privacy*: In what way could robotics, in all its applications, affect human functions, capabilities, and rights related to human data protection and privacy issues? On the side of positive rights, how could these technologies suggest a further definition of the concepts of human liberty, dignity, and identity?
- *Preservation of human identity and transhumans*: In the field of bionics and robotics prosthetics, what is the border between restoration and enhancement? What is the human identity that robotics implants should preserve?
- *Liability and responsibility issues*: Who is responsible for the possible malfunctions and damages by autonomous robots to humans and/or property? (issues in the domain of AI and law. This subject also implies the identification of new possible additions to the definition of personhood and agentive capacities and the responsibilities discharged regarding robots).
- *Psychological effects:* The effects of personal robots on human logical and emotional structures and on human relationships (psychology, sociology, HRI, neurosciences, and law).
- *Cost-benefit analysis*: The comparison between robotics applications and all other possible alternatives, evaluating

whether or not to add the ethical element among the variables listed in the cost-benefit analysis. This subject also comprises the issues of human replaced by robot, job displacement, and analysis of new educational requirements and professional qualification of the human workforce and operators (economy, public policy, engineering economics, and corporate social responsibility).

Initially, the fields of robotics that are going to raise more urgent and entangled ethical issues are biorobotics and robotics military applications. The reason is that these two sectors have, with their research and prototypal applications, immediately and directly intervened in the core of the existing corpus of ethical principles, regulations, and laws related to the most sensitive issues of human life. In less dramatic but not less-important instances, assistive and educational robotics has also given rise to some ethical concerns in the field of HRI.

Because of the rich and complex debate in roboethics and the far-reaching scenarios that could develop over the following decades, this article (with the partnership of the aforementioned European projects) adopted a triaging work methodology, analyzing the issues that had the following elements:

- *Novelty*: Issues that have never been coped with; the absentia legis and the lack of regulations, in many cases (bionics and military robotics), underpin a severe responsibility gap.
- *Emerging*: Issues arising in a nonprogrammed way, as the prototypal robotics products are the results of the drive of different instances: research and business.
- *Complexity*: Issues lying in the crossroads of several disciplines (robotics, AI, moral philosophy, psychology, anthropology, and law).
- *Social pervasiveness*: Issues related to current and yet-to-be-released robotics products.

Ethics in Complex Technological Societies

For the purpose of this study, we will provide the various meanings of the term *ethics* within the context of modern debate. Traditionally, ethics is the philosophical or theological subject that studies human behavior and assessment criteria for human behaviors and choices. Modern ethics have developed from various points of view along classical philosophy. In the last decades, in our highly complex societies characterized by technoscientific development, the attribution of moral and practical responsibility is becoming more and more difficult with regard to unintentional or collateral consequences of actions and operations that are produced by unidentifiable decisions (group decisions, complex administrative structure, distribution of responsibilities, and computerized operations). Often, one finds it impossible to attribute final responsibility to a single person or to a defined social entity. In the absence of a definition and a precise analysis of the responsibility chain, technologically advanced societies have shifted the issue onto the concept of risk assessment, thereby attributing value to the damage produced by entities that are seemingly devoid of responsibility. In our case, the question is: Who is responsible for any damage that may be caused by an autonomous robot? Is it the designer, manufacturer, programmer, or final user? Often, it will be difficult to obtain easy answers to this question.

From the viewpoint of ethical theories, we observe that individuals possess the ethics of common sense, which provide them with the moral guidelines for decision making, from the small ones to important life decisions. This affects the day-to-day life as well as our actions in our close-knit social relations. Individuals often adopt different ethical theories from time to time: utilitarianists for some decisions, in other situations, or moments generally go by the terms of their upbringing, customs, tradition, and religion (descriptive ethics). These seem to be enough for them.

However, this line of thought reaches its limits when, as social figures or in our profession, we are faced with complex problems, in which our actions may have multiple consequences that are difficult to follow and predict; and when our common sense is faced with the problems we've never dealt with before—for example, bioethical dilemmas. In these situations, the ethics of common sense leads to various paradoxes: we find ourselves without any conceptual resources and in the difficult position of having to pass judgment. In these cases, we need a logical-critical set of ethics (critical ethics) that 1) reveals the implicit and, perhaps, never uncovered assumptions in our ethics of common sense and in our outdated ethical theoretical base and 2) analyses the reasons, the pros and cons, and their origin. Inevitably—we may not even realize it—we resort to prescriptive ethics and ethical theories.

In practice, when faced with general and complex topics, we may refer to the fundamental, relevant values involved in our dilemmas; we may adhere to the more updated morality applied to issues close to ours; or we try to step up from the universally shared prescriptions toward new ethical frontiers. Prescriptive ethical theory, which develops and justifies the principles of moral actions, refers to ethical theories and related guidelines. These in turn represent the general ideas that contain ethical principles to reach an internal and systematic coherence. In defense of the ethical principle, the ethical philosophers either implicitly or explicitly refer to an ethical theory.

In the 20th century, because of the dissatisfaction raised by the issues that are deemed to be the limits to traditional ethical theory—utilitarianism and deontologism inspired by Kant—ethics became fragmented into several forms: rights ethics, virtue ethics, feminist ethics, and applied ethics [9].

For instance, according to virtue ethics—which have less to do with single actions, but more to do with different life styles and ways of life—the most important moral obligation is our personal relationship with the action of making or using robots and the analysis of the effect of robots with respect to the concept of good and happiness: Which robots could contribute most to the happiness to the full and complete quality of life of human beings? [10]. In

Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

another instance, considering the need to account for values that are part of a collective conscience, the so-called rights ethics sees human rights to be the most relevant elements in common for a system of cultural and moral pluralism and as being reasonably deemed to be the final manifestation of universal ethics.

One of the challenges to traditional ethical theories by some thinkers regarded the limitation of moral considerations to the problems of humans and their relationship in human society, regardless of other living organisms and environment. Therefore, ethical theories and the new applied ethical theories, in particular, were called upon to include nonhuman entities in their analysis (animal ethics, environmental ethics, and planetary ethics known as the *Gaia theory*) as well as human products (bioculture, computer ethics, and roboethics).

The various applied ethical theories are, in turn, connected and intertwined with other disciplines, including law, sociology (descriptive ethics), economics, and various scientific fields. One of the central themes of applied ethics is the concept of responsibility, which is a moral notion. Legal responsibility determines the rules of the relevant prescriptive ethics, meaning the group of commands and prohibitions adopted by a society or group and that also defines professional ethics. For the purpose of this article (to analyze the human–robot relationship), we shall consider the two main meanings of the term responsibility:

1) the analysis of the identity of the agent of the cause of certain actions and their effects (utilitarianism or consequentialism or teleologism)
2) an expression of motivations that leads an agent to act in a certain way (deontological ethics or Kantian ethics), according to which the individual assesses the consequences of his or her actions.

In the last century, we know that the questions "What authority and what set of moral rules am I obliged to be accountable to? State law? God?" forced many answers (utilitarian or deontological) to a crisis point. Often, the moral response of the individual as well as the moral evolution of our society have led to an opposition to the responsibility toward the nation, church, or traditional roles of social institutions (see Max Weber's ethics of intentions or ethics of individual conscience).

Some serious events related to World War II changed the notion of responsibility and differentiation of roles (i.e., engineers deal with engineering, doctors with medicine, soldiers obey orders from superiors, etc.) and also some famous legal cases, e.g., the Adolph Eichmann trial, during which he defended himself by stating that he was obeying the orders under his administrative responsibilities and that he did not decide on, nor was he aware, of the entire project of Jewish extermination or the debate that ensued after Hiroshima and Nagasaki 1945.

Furthermore, there are different cases in which our personal moral clashes with those adopted and imposed by the society we live in and whose law we live under. Such is the case, for example, of the death penalty (which is in force in various states in the United States, although contested by internal currents), animal rights, abortion, or euthanasia. In these cases, the implicit moral philosophy in state laws does not coincide with the feelings of many groups of the relevant societies, which leads to this sort of conflict.

Therefore, we observe that, in contemporary societies, the notion of responsibility is not limited to the moral consideration of actions of the agent or cause and deals with the needed conformity of the action with a group of duties (therefore, the analysis of consequences is less decisive). In today's society, just the elements of complexity and technological ruling determines the aspect known as *heterogenesis of ends*, according to which our actions may have consequences that are extremely difficult to estimate, which may even be opposite to our intentions. According to Morin's ecology of action [11], once the action departs from the individual, it lives a life of its own, and combines itself with the environmental conditions (social models, actions, and reactions of other agents), and the final result is beyond the agent's predictive abilities.

Faced with this vacancy in the attribution of individual responsibility, some researchers attempted to identify collective shared responsibilities, were it to be impossible or vain to identify an individual responsibility. In the case of scientific research, science and technology studies (S&TS) expert Renè von Schomberg proposes to adopt an assessment system based on foresight and knowledge (foresight and knowledge assessment). The author sustains that, because the definition of responsibility is considerably more arduous to define in scientific fields, due to the unintentional consequences, uncertainty, or ignorance of results, instead of identifying the ethical responsibilities post hoc, it is necessary to establish the ethics of the overlap of knowledge between different areas beforehand (synergy: scientists, politicians, etc.), because the quality of knowledge shall determine the ethical value of the applications that will follow.

At the same time, one must constantly ensure that maximum precision of predictions to identify both the wholesomeness of research and relevant applications as well as the potential ethical problems [12].

Other authors have emphasized the need to avoid the overlap between ethical problems and technical solutions: among the latter, the expert of computer ethics, Abbe Mowshowitz [22] states that

> the seemingly eternal social problems are real enough, but to look for their cause in technique or autonomous technology is both mistaken and harmful. We should not blame technology for human failures. (..) Autonomous technology contributes to the belief in technological determinism, i.e., reinforces belief in the inability of people to make significant choices in their lives. It directs attention away from wielders of power to systems of reified collectivities. The law is smarter than the social sciences—it defines the corporation, for example, as a fictional person for purposes

of assigning responsibility and does not absolve key actors of their responsibilities. Institutions should be seen as convenient fictions that help explain individual decision and action (..) Only the actions of human beings can be alienating or dehumanizing. Reification of technology allows for an illegitimate transference of responsibility from persons to a fictional social construct, and at the same time, impedes our ability to come to grips with the very real ethical challenges posed by the uses of technology.

In conclusion, we observe that, in roboethics, the definition of moral responsibility and the resulting notion of liability—which is central in human–robot relationship—could differ according to the philosophical assumptions which, knowingly or unknowingly, have been adopted.

Roboethics or Robot Ethics?

Roboethics was originally conceived as human-responsibility ethics. The roboticists and ethicists that contributed to their creation highlighted the following aspects:

- problems regarding robot autonomy
- problems related to warfare applications
- problems in human–robot relations (dependence, privacy, robot appearance, and potential confusion between natural and artificial)
- digital divide (for nations, genders, and ages)
- ethical dimension of technology.

According to Veruggio [13],

> roboethics is an applied ethics whose objective is to develop scientific/cultural/technical tools that can be shared by different social groups and beliefs. These tools aim to promote and encourage the development of robotics for the advancement of human society and individuals and to help preventing its misuse against humankind.

This and similar definitions imply that robotics and its applications are subject to moral judgment and human intervention. According to this perspective, roboethics is not artificial ethics and not even the regulatory system of dependability and safety. Roboethics indicated that the individual and society may intervene upstream on the direction of robotics and its products. The individual may, for example, limit the use of robots, according to a wider precautionary principle, in the absence of the necessary safety precautions when missed for humans.

As researchers, individuals may refuse to design robots that are deemed harmful or hazardous; as professional bodies, they may discuss and decide upon an appropriate professional ethic.

As already discussed, roboethics has adopted the principles established by the Charter of Human Rights and the Lisbon Treaty. However, these certainly do not satisfy the realm or depth of the ethical debate. If last century extended the ethical realm to an increasing number of elements, including animals and our planet among those entitled to rights, we witness an increasing application rights ethics to even broader categories, thereby crossing over the human/organic barrier.

In the definition of roboethics, as developed by some authors [6], [14], the agreement with the thesis of American moral and political philosopher John Rawl and to his reflective equilibrium is explicit, according to which the charters and treaties, although advanced in an attempt to associate the highest number of participants and commendable in an attempt to widen the promotion of rights to a greater number of members and functions as possible, do not satisfy the possibility of positive progress, nor can they provide for the entire range of criminal uses of robots.

During the past six years of discussion on roboethics, various positions have emerged regarding the ELS issues in robotics. Just as some authors have continued to use the term roboethics, others have used robot ethics. These two terms don't always indicate different notions. However, robot ethics has been often used to indicate 1) the artificial ethics of robots or the morality of robots and 2) the group of prescriptions, rules, and regulations regarding robot safety and dependability. Although the term robot ethics may be used in the latter sense, because roboethics also studies issues of dependability and safety, the difference between roboethics and robot ethics is more complex.

Some authors see robot ethics as an artificial morality and the robots as autonomous agents. Their theses in this context are different, but they can be linked to this cluster of arguments.

The first states that sophisticated autonomous robots, because they are equipped with intelligence and a certain kind of conscience, they include an ethical system that can learn and evolve or are able to decide between good or evil. According to this position, morality and immorality constitute a gradual continuum: such is the case for children, which, in our society, are not deemed to be fully responsible for their actions or the disabled [15], [16]. As they are quasi-moral agents, robots are also subject to ethical behavior, equipped with some rights of their own.

The second position in robotics morality states that robots, as devoid of emotion and passion, and as they are equipped with rational behavior (because that is how they are programmed), could be more ethical than human beings [17]. These two positions have one aspect in common: they state that, when faced with the complexity of robot technology, it is difficult to assign responsibility for any possible damage they may cause; or when faced with irrational and complex situations (war theaters and other circumstances where an immediate answer—such as a gut reaction—is required), the placing of trust in an automated morality is the only solution. Thus, the focus has shifted from the issue of human ethics and has moved onto operational artificial ethics, which therefore avoids the debate begun by Sanremo [5] and that was essentially focused on the issue: when is it right to limit the autonomy of robots as final products?

When one tries to shift an ethical problem onto a technical solution, there are some concepts that fall into the background, such as 1) important underlying assumptions about the adopted ethical theories; 2) the relevant glossary, meaning an ambiguous use of terminology. In our case, here the keyword is *machine learning*.

Although searching for new solutions is more than laudable and auspicious—for instances, when faced with the horror of war and above all those of this century—coherence with the philosophy of law imposes that every new form of ethics cannot demean the rights acquired by the previous ones: "Any new ethics must deal with the same substance as the old role responsibility ethics, namely, with values and norms that restrict or delimit human action and thus enable or guide traditional decision making" [12]. This statement means that roboethics and robot ethics should be coherent with the shared sets of rules.

Overall, every thesis in favor of robotics morality per se is based on the assumption of predictability (or rationality and foreseeability) of the behavior of an autonomous robot. A robotic morality provides, in the conception of its authors, that the robot shall behave ethically according to what it has been programmed with and what it has learned. In a word, it cannot deviate from its incorporated laws: unlike a human being that, although supposedly brought up to be good, may decide to do evil unexpectedly, a robot will always decide ethically because it cannot behave otherwise.

Here, the critical analysis should focus on two critical elements: 1) the kind of ethics that the robot expresses and 2) the unpredictability issue in leaning machines. The first question implies, in our view, that no robot ethics can avoid the needed, deep and broad debate on the human ethical principles grounding roboethics and their enforcement. The second critical element is that, according various experts [18], [19], "programmers, manufacturers, and users may not be in the position to predict what a learning robot will do in normal operating environments and to select an appropriate course of action on the basis of this prediction" [3]. This means that not only it would be advisable that learning robots should be self-evidently distinguishable from nonlearning robots, but that in the former case, the learning process be made transparent for the robot user.

An ethically accurate analysis of the bonds and limits of the potential sphere of action of a robotics morality should indeed take into careful account the epistemic and logic-hidden faults related to both ethical theories and technical constraints. The fact that the current development of non-segregated robotic systems does not yet allow for a precise modeling of every possible environmental factor or for a final definition of normalcy in rich operating conditions should be the basis for a more prudent discourse on intelligent robots behaving ethically.

Moreover, a careful assessment of the responsibility issues in attributing to robot rights, which belong only to humans, and the analysis of the problems encountered in assessing the environmental factors affecting robotics behavior, would advise to adopt the enlarged version of the precautionary principle, limiting robotics autonomy when all the environmental factors are not precisely assessed.

Conclusions: An Open Debate

From the manifold and depth of the considerations around roboethics over the last six years and considering the new ethical issues involved that have never been broached, it appears as if, beyond the different point of views, one major side question is posed, which would need a more general answer: Which direction robotics is going to take? And which should it take? The hope of the many, whose consideration of the importance of robots in society, goes hand in hand with ethical related concerns, is that a promising alliance between robotics and the field of science and technology studies should also be swiftly established.

It has been widely recognized—although not very often practically accepted—that institutional practices in science and technology are tacitly shaped and framed by deeper social values and interests [20]. These include

- important political–economic relationship to science and technology
- shifting from science as independent republic to science as cooperating in innovation and in the knowledge economy
- impacts of the increasing commercialization of science in particular areas affecting public trust
- loss of credibility and senses of unease in the general public [21].

The researches in science and society developed by social scientist Sheila Jasanoff and collaborators have sustained that technoscientific knowledge stabilizes in society through a complex and articulated process of negotiation and then very seldom experts' opinions are exempt from uncovered assumptions.

It should be clear to all the parties involved in the process of molding roboethics that this free, open, and relatively untroubled debate is possible, because until now no dramatic incident in the field occurred. And, we hope it never will. Robotics research is driven by future scientific visions, and it so should be. Limitations of the freedom of scientific research should be very carefully discussed and almost never imposed. Limitations should be painstakingly decided on for marketed robotics applications.

However, some from the robotics industry have already expressed their concern that, at the first sense of unease from some social groups toward new robotics applications, some limits will be decided on. Increasingly, if a dramatic incident was to occur, the extraordinary character of the situation will impose severe constraints. We have already witnessed those occurrences in bioethics (stem cells, etc.). Among the recommendations resulting from the analysis of ELS issues in robotics developed in the frame of CARE, there is a need to "avoid that ELS issues in robotics could become a barrier to further progress of our field."

If some incidents were to occur, the outdoor process of discussion will be channeled in the indoor environment of the experts committee, relegating the other parties (roboticists,

IEEE Robotics & Automation MAGAZINE | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

ethicists, stakeholders, and society as a whole) to the role of concerned observers.

For these reasons, we feel that all people concerned with roboethics should take vantage from this fortunate window of fresh and free debate, defining with careful consideration and wise temperance of language the general ethical assessments and rules for future robotics.

References

[1] B. Siciliano and O. Khatib, *Springer Handbook of Robotics*. Berlin Heidelberg: Springer-Verlag, 2008.

[2] C. Breazeal, "Emotion and sociable humanoid robots," *Int. J. Human Comput. Stud*, vol. 59, no. 1–2, pp. 119–115, July 2003.

[3] G. Tamburrini and D. Marino, "Learning robots and human responsibility," *Int. Rev. Inform. Ethics*, vol. 6, pp. 46–52, Dec. 2006.

[4] G. Veruggio, "Io robotico," *Le Scienze (*Italian Edition *of Scientific American)*, no. 434, pp. 40–47, Oct. 2004.

[5] G. Veruggio, (2005, Apr. 18). The birth of roboethics. presented at Proc. IEEE Int. Conf. Robotics and Automation Workshop on Roboethics (ICRA 2005), Barcelona [Online]. Available: http://www.roboethics.org/icra2005/veruggio.pdf

[6] G. Veruggio, (2006, Dec. 4–6). The Euron Roboethics Atelier. For the Euron Roboethics Roadmap. presented at Proc. 6th IEEE-RAS Int. Conf. Humanoid Robots, Genoa, Italy [Online]. Available: http://www.euron.org/activities/projects/roboethics

[7] ETHICBOTS project. Emerging Technoethics of Human Interaction with Communication, Bionic, and robOTic systems. (2005, Nov. 1–2008, Apr. 30). Coordination action, FP6—Science and society [Online]. Available: http://ethicbots.na.infn.it/

[8] CARE. Coordination Action for Robotics in Europe. The strategic research agenda. *European Robotics Technology Platform*, [Online]. Available: http://www.robotics-platform.eu/cms/index.php?idcat=26

[9] A. Fabris, *Quale etica per la robotica?* in *Ethicbots, Etica e Robotica*, (Teoria, Rivista di filosofia). Pisa: Edizioni ETS, 2007, pp. 7–17.

[10] M. Nussbaum and A. Sen, *The Quality of Life*. Oxford: Clarendon, 1993.

[11] E. Morin, *La Methode*. Ethique: Editions de Seuil, 2004.

[12] R. von Schomberg, "From the ethics of technology towards an ethics of knowledge policy: Implications for robotics," *AI Soc.*, vol. 22, no. 3, pp. 331–345, 2007.

[13] G. Veruggio and F. Operto, "Ethical, legal and societal issues in the strategic research agenda of the coordination action for robotics in Europe," in *Proc. Workshop on Roboethics (ICRA09)*, Kobe, Japan, May 17, 2009.

[14] G. Tamburrini, "Robot ethics: A view from the philosophy of science," in *Ethics and Robotics*, R. Capurro and M. Nagenborg, Eds. Heidelberg: IOS Press, 2009, pp. 11–23.

[15] P. M. Asaro, "What should we want from a robot ethic?," in *Ethics and Robotics*, M. Capurro and M. Nagenborg, Eds. Heidelberg: IOS Press, 2009, pp. 1–11.

[16] W. Wallach and C. Allen, *Moral Machines: Teaching Robots Right from Wrong*. Oxford, New York: Oxford Univ. Press, 2009.

[17] R. Arkin, *Governing Lethal Behavior in Autonomous Robots*, Boca Raton: CRC Press, 2009.

[18] J. Altmann, "Preventive arms control for uninhabited military vehicles," in *Ethics and Robotics*, R. Capurro and M. Nagenborg, Eds. Heidelberg: IOS Press, 2009, pp. 69–83.

[19] C. Tamburrini, "AI and Popper's solution to the problem of induction," in *Karl Popper: A Centennial Assessment*, vol. 2, *Metaphysics and Epistemology*, I. Jarvie, K. Milford, and D. Miller, Eds. London: Ashgate, pp. 265–282.

[20] M. Callon, M. E. Gonçalves, S. Jasanoff, M. Jepsen, P. B. Joly, Z. Konopasek, S. May, C. Neubauer, A. Rip, K. Siune, A. Stirling, and M. Tallacchini, "Taking European knowledge society seriously," Report of the Expert Group on Science and Governance to the Science, European Commission, Directorate-General for Research, Science, Economy and Society Editor, Brussels, Report no. EUR 22700, 2007.

[21] M. Polanyi, *Personal Knowledge–Towards a Postcritical Philosophy*, London: Routledge, 1958.

[22] A. Mowshowitz, "Technology as excuse for questionable ethics," *AI Soc.*, vol. 22, no. 3, pp. 271–282, 2008.

[23] D. Bloor, *Knowledge and Social Imagery*, London: Routledge, 1976.

[24] G. Veruggio and K. Abney, "Roboethics: The applied ethics for a new science," in *Robot Ethics: The Ethical and Social Implications of Robotics*, P. Lin, K. Abney, and G. Bekey, Eds. Cambridge, MA: MIT Press, to be published.

[25] G. Veruggio and F. Operto, "Roboethics: Social and ethical implications of robotics," in *Handbook of Robotics*. Heidelberg: Springer-Verlag, 2008, ch. 64.

[26] J. H. Moor, "What is computer ethics?" in *Computers and Ethics*, T. W. Bynum, Ed. Oxford, U.K.: Blackwell, 1985, pp. 266–275, Also in *Metaphilosophy,* vol. 16, no. 4, Oct. 1985.

[27] J. Weber, "Robotic warfare, human rights and the rhetoric of ethical machines," in *Ethics and Robotics*, R. Capurro and M. Nagenborg, Eds. Heidelberg: IOS Press, 2009, pp. 83–105.

[28] L. Floridi, "Information ethics: Its nature and scope," in *Information Technology and Moral Philosophy*, J. van den Hoven and J. Weckert, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2008, pp. 40–65.

[29] M. Nagenborg, R. Capurro, J. Weber, and Ch. Pingel, "Ethical regulations on robotics in Europe," *AI Soc*, vol. 22, no. 3, pp. 349–366, 2007.

[30] N. Sharkey, "The ethical frontiers of robotics," *Science*, vol. 322, no. 5909, pp. 1800–1801, 2008.

[31] S. Gill and G. Tamburrini, "Guest editorial," *AI Soc.*, vol. 3, no. 22, pp. 265–270, 2007.

[32] S. Jasanoff, *The Fifth Branch: Science Advisers as Policymakers*. Cambridge, MA: Harvard Univ. Press, 1990.

[33] M. Santono, D. Marino, and G. Tamburrini, "Learning robots interacting with humans: From epistemic risk to responsibility," *AI Soc.*, vol. 3, no. 22, pp. 301–314, 2007.

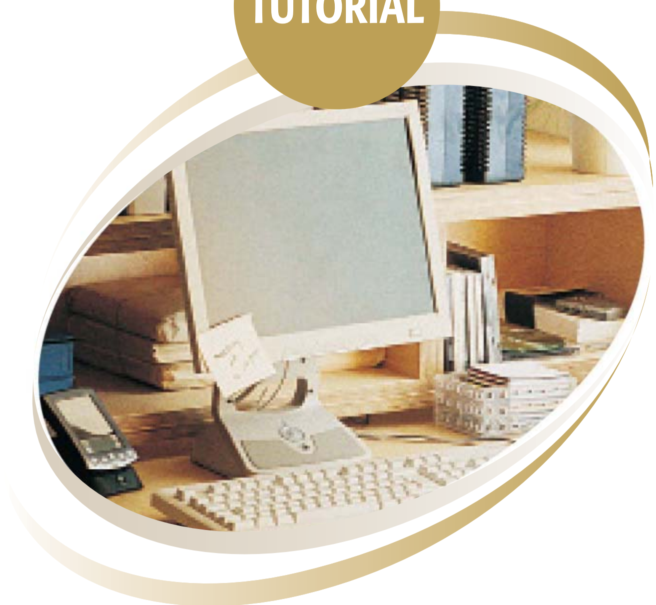
[34] T. Bynum, W. Maner, and J. Fodor, Eds. *Teaching Computer Ethics*. New Haven: Research Center on Computing & Society, 1993.

[35] T. W. Bynum, "Norbert Wiener and the rise of information ethics," in *Information Technology and Moral Philosophy*, J. van den Hoven and J. Weckert, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2008, pp. 8–25.

Fiorella Operto, Scuola di Robotica, P.O. Box 4124, P.za Monastero 4, 16149 Genova, Italy. E-mail: operto@scuoladirobotica.it.

Robotics & Automation Magazine | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

TUTORIAL



Motion Planning

Part I: The Essentials

By Steven M. LaValle

This is the first installment of a two-part tutorial. The goal of the first part is to give the reader a basic understanding of the technical issues and types of approaches in solving the basic path-planning or obstacle-avoidance problem. The second installment will cover more advanced issues, including feedback, differential constraints, and uncertainty. Note that this is a brief tutorial rather than a comprehensive survey of methods. For the latter, consult some of the recent textbooks [4], [9].

Motion planning involves getting a robot to automatically determine how to move while avoiding collisions with obstacles. Its original formulation, called *the piano mover's problem*, is imagined as determining how to move a complicated piece of furniture through a cluttered house. Have you ever argued about how to move a sofa up a stairwell? It has been clear for several decades that getting robots to reason geometrically about their environments and synthesize such plans is a fundamental difficulty that recurs all over robotics.

Digital Object Identifier 10.1109/MRA.2011.940276
Date of publication: 14 April 2011

The stages of motion-planning development are parellel to those of an integral calculus: 1) The integration problem was clearly identified and defined; 2) perfect, exact solutions were developed for many classes of functions; and 3) since these were limited to a small subset of functions that people care about, numerical integration methods were developed with great success in practice. The similar stages of motion planning were as follows: 1) it was clearly defined in the 1970s; 2) the 1980s saw the development of perfect, combinatorial solutions, which are ideal in some settings, but not practical in most; and 3) the 1990s brought sampling-based methods that are not as elegant but offer practical solutions to modern industrial-grade problems. Over the past decade, motion-planning algorithms have been widely used in robotics and automation and have furthermore found applications well beyond, including the fields of virtual prototyping and computational biology.

Problem Formulation

Let $\mathcal{W}$ denote the world that contains a robot and obstacles. For a two-dimensional (2-D) world, $\mathcal{W} = \mathbb{R}^2$ and $\mathcal{O} \subset \mathcal{W}$ is the obstacle region, which has a piecewise-linear

IEEE Robotics & Automation Magazine
Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

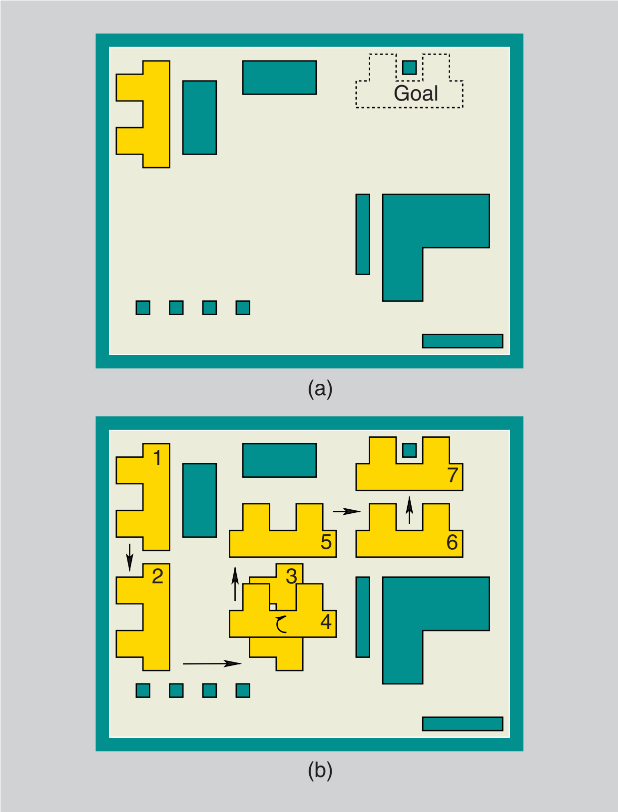


Figure 1. A 2-D example of basic path planning.

(polygonal) boundary. (The complement $\mathcal{W}/\mathcal{O}$ is assumed to be a bounded open set.) The robot is a rigid polygon that can move through the world but must avoid touching the obstacle region. For a three-dimensional (3-D) world, the only differences are that $\mathcal{W} = \mathbb{R}^3$, and $\mathcal{O}$ and the robot are defined with polyhedra instead of polygons. Motion-planning formulations extend well beyond the rigid polygons and polyhedra, but such extensions are left to the "Direct Extensions" section and the second part of this tutorial.

The basic path-planning problem is informally summarized as follows: given an initial placement of the robot, compute how to gradually move it into a desired goal placement so that it never touches the obstacle region. See Figures 1 and 2 for examples.

Consider the task in terms of algorithm inputs and outputs.

- *Inputs*: An initial placement of the robot, a desired goal placement, and a geometric description of the robot and obstacle region.
- *Outputs*: A precise description of how to move the robot gradually from its initial placement to the goal placement while never touching the obstacle region.

The output description will be a path through a set of all intermediate transformations of the robot from start to finish.

Living in C-Space

Although the motion-planning problem is described in the world, it really lives in another space: the set of all rigid-body transformations that can be applied to the robot is called the *configuration space* or *C-space*. Finding a solution leads to computing a path through the part of the C-space that avoids robot-obstacle collisions.

A rigid body may translate and rotate. Most people are much more familiar with performing one transformation to place a body into a scene rather than thinking about all transformations. The notion of configuration space was the key insight to Lagrangian mechanics of rigid bodies [1], as it allowed dynamics to be expressed using the precise degrees of freedom of a body. The idea was introduced to motion planning by Lozano-Perez [12] and Udupa [17]. The C-space in physics and control theory is usually called a *Lie* (pronounced Lee) *group*. In this context, which is much more widely studied than motion planning, the C-space is considered as a differentiable manifold, which leads to considerable technical and notational hurdles. The C-space used in motion planning requires no calculus; therefore, it is described as a *topological manifold*, which is fortunately much simpler to define and manipulate. The definition of an n-dimensional (topological) manifold $\mathcal{C}$ is a subset of $\mathbb{R}^m$ for $n \leq m$, such that every $q \in \mathcal{C}$ is contained in at least one open subset of $\mathcal{C}$ (pick a small one) that is homeomorphic. (Homeomorphic means that for an open set, say O, there exists a continuous, bijective function $f : O \to \mathbb{R}^n$ for which the inverse f^{-1} is also continuous to $\mathbb{R}^n$.) The intuition is that, in the local vicinity of every q, a manifold behaves like $\mathbb{R}^n$. It is a nicely behaved surface. The existence of sharp corners does not even matter;

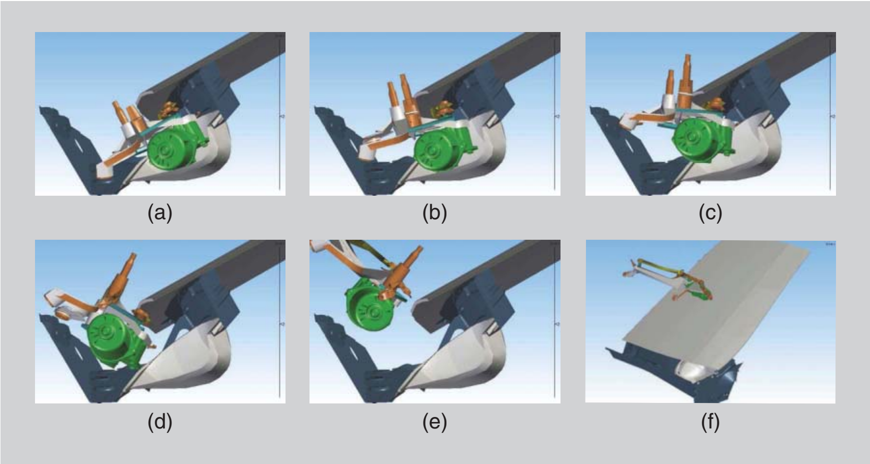


Figure 2. A 3-D automotive assembly task that involves inserting or removing a windshield wiper motor from a car body cavity. This problem was solved for clients using the path-planning software of Kineo CAM.

IEEE Robotics & Automation Magazine | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

however, branching or the locally changing dimensions is not allowed (Figure 3).

We now take a look at the C-spaces that commonly arise in planning. Consider a 2-D world. Let $\mathcal{A} \subset \mathbb{R}^2$ denote a polygonal robot. It could, for example, be all points inside of a triangle defined by vertices $(-1,0)$, $(1,0)$, and $(0,1)$. We could rotate the robot counterclockwise by any $\theta \in [0, 2\pi)$ and then translate it by any $x_t \in \mathbb{R}$ in the X direction and any $x_t \in \mathbb{R}$ in the Y direction. This allows for any possible position and orientation, and every x_t, y_t, θ combination leads to a unique robot placement. Let $q = (x_t, y_t, \theta)$ be called the *configuration*. A point $(x, y) \in \mathcal{A}$ would then appear at some $(x', y') \in \mathcal{W}$ (in the world) given by

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta & x_t \\ \sin\theta & \cos\theta & y_t \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \quad (1)$$

which uses a standard 3 by 3 homogeneous transformation matrix. The upper left 2 by 2 block is just a rotation matrix.

The set of all configurations $q = (x_t, y_t, \theta)$ is clearly a subset of $\mathbb{R}^3$, but to define the C-space, we must take into account that $\theta \pm 2\pi$ yields equivalent rotations. We write that $C = \mathbb{R}^2 \times S^1$, in which S^1 denotes a circle in the topological sense and accounts for θ (the circle is obtained by gluing 0 and π together). The C-space $\mathcal{C}$ is a 3-D manifold, and each element is nicely described as $q = (x_t, y_t, \theta)$. Remembering that θ wraps around at 2π is crucial to motion planning; otherwise, an artificial barrier or redundant exploration will be introduced. If the robot is not allowed to rotate, then we obtain the translation-only case and $C = \mathbb{R}^2$ with $q = (x_t, y_t)$.

For the 3-D world, the concepts mostly extend as you might expect. Three translation parameters x_t, y_t, z_t appear, and a translation-only robot then has a C-space $\mathcal{C} = \mathbb{R}^3$ with $q = (x_t, y_t, z_t)$. However, the set of 3-D rotations turns out to be 3-D manifold all by itself, and it is not as simple as a circle or sphere topologically. The best way to see its structure is to use quaternions to represent rotations. Since this a brief tutorial, only the essence is given here, and quaternion algebra is avoided here as it is not critical to motion planning. Every 3-D rotation can be expressed as a rotation by an angle $\theta \in [0, 2\pi)$ about some fixed axis that passes through the origin. Let this axis be described by some unit vector $v = (v_1, v_2, v_3)$. This already makes it appear that there is a sphere of possible axes and then a circle of possible angles at each place on the sphere. This collection of circles glued together around the sphere is called *Hopf fibration*. Now there is another trouble. Just as 0 and 2π were equivalent in the 2-D case; for the 3-D case, we have that v and θ to produce the same rotation as $-v$ and $2\pi - \theta$. A convenient way to handle this is to define $h = (a, b, c, d)$ and assign $a = \cos(\theta/2)$, $b = v_1 \sin(\theta/2)$, $c = v_2 \sin(\theta/2)$, and $d = v_3 \sin(\theta/2)$. Note that $a^2 + b^2 + c^2 + d^2 = 1$, meaning that h lies on a unit sphere. Furthermore, h and $-h$ are equivalent rotations. The C-space for the set of all 3-D rotations is therefore nicely visualized as a 3-D sphere, a subset of $\mathbb{R}^4$ in which opposite (called *antipodal*) points are the same. This means that, to get the set of all rotations, we can stay in the upper hemisphere ($a \geq 0$), but must be careful at $a = 0$, because opposite points on this equator are the same. The technical term for the resulting space is *real projective three space*, denoted $\mathbb{RP}^3$. For the case of a 3-D robot that can translate or rotate, we obtain $C = \mathbb{R}^3 \times \mathbb{RP}^3$, which is a six-dimensional manifold. We can represent the configuration as $(x_t, y_t, z_t, a, b, c, d)$ while enforcing that $a^2 + b^2 + c^2 + d^2 = 1$. The use of quaternions means that the set of all 3 by 3 rotation matrices is parameterized by $a, b, \mathcal{C}$, and d:

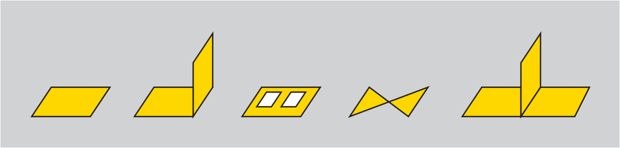


Figure 3. The first three are manifold, because they locally look like $\mathbb{R}^2$; the last two are not because at some points the dimension changes or branching occurs.

$$\begin{pmatrix} 2(a^2+b^2)-1 & 2(bc-ad) & 2(bd+ac) \\ 2(bc+ad) & 2(a^2+c^2)-1 & 2(cd-ab) \\ 2(bd-ac) & 2(cd+ab) & 2(a^2+d^2)-1 \end{pmatrix}. \quad (2)$$

With different possible parameterizations of rotations, for 2-D or 3-D worlds, it is important to realize that if two points are close under one representation, they might be far under another. Furthermore, if there are singularities in the parameterization mapping (e.g., yaw–pitch–roll representation), the C-space might not even represent the same manifold as the set of all rotations.

Now that different possibilities for $\mathcal{C}$ have been presented, consider the parts of $\mathcal{C}$ that are prohibited due to collision. Let $\mathcal{A}(q) \subset \mathcal{W}$ denote a closed set of points in the world occupied by the robot $\mathcal{A}$ when it transformed to configuration q. A configuration $q \in \mathcal{C}$ places the robot into collision if and only if $\mathcal{A}(q) \cap \mathcal{O} \neq \emptyset$ (the robot and obstacle are attempting to occupy at least one common point in $\mathcal{W}$). The set of all noncolliding configurations is often called the free space and is defined as

$$\mathcal{C}_{\text{free}} = \{q \in \mathcal{C} \mid \mathcal{A}(q) \cap \mathcal{O} = \emptyset\}. \quad (3)$$

The complement is called the *obstacle region in C-space*: $\mathcal{C}_{\text{obs}} = \mathcal{C}/\mathcal{C}_{\text{free}}$.

The problem statement given in the "Problem Formulation" section seemed somewhat informal; however, using the C-space, the basic path-planning problem can be precisely defined: given a robot description $\mathcal{A}$, an obstacle description $\mathcal{O}$, a C-space $\mathcal{C}$, an initial configuration $q_I \in \mathcal{C}$, and a goal configuration q_G, compute a continuous path $\tau : [0,1] \to \mathcal{C}_{\text{free}}$ with $\tau(0) = q_I$ and $\tau(1) = q_G$ (Figure 4). A typical way to express τ is a sequence of line segments, which ignores the particular parameter $s \in [0, 1]$, but is good enough for motion-planning results. Note that the path must be continuous; otherwise, the robot would appear to teleport from one place to another, which is obviously cheating. Gradual motions through $\mathcal{C}$ make the robot move gradually through $\mathcal{W}$.

Combinatorial Planning

Although the motion-planning problem is in the continuous C-space, its computation is discrete. Therefore, if we want an algorithmic solution, we need a way to discretize the problem. This has led to two main schools of thought: 1) combinatorial planning, which thrived in the 1980s, constructs structures in the C-space that discretely and completely capture all information needed to perform planning and 2) sampling-based planning, developed mainly across the 1990s, uses collision-detection algorithms to probe and incrementally search the C-space for a solution rather than completely characterizing all of the $\mathcal{C}_{\text{free}}$ structure. The second approach is most widely used in practice; however, the first one is far superior in many instances. Therefore, it is worth to study both.

To illustrate the philosophy of combinatorial planning, consider the case in which $\mathcal{W} = \mathbb{R}^2$ and contains a point robot ($\mathcal{A} = \{(0, 0)\}$) that cannot rotate. In this case, $\mathcal{C} = \mathbb{R}^2$, and the task is simply to connect the dots in the plane with a curve that avoids the obstacles [Figure 5(a)].

Here is a simple technique that contains all the essential ingredients of combinatorial planning. All the methods first compute a road map, which is a graph in which each vertex is a configuration in $\mathcal{C}_{\text{free}}$, and each edge is a simple path through $\mathcal{C}_{\text{free}}$ that connects a pair of vertices. Here is one way to achieve this:

1) Decompose $\mathcal{C}_{\text{free}}$ into trapezoids with vertical side segments. Figure 5(b) shows the result. From each polygon vertex, an attempt is made to shoot rays upward and downward. Each ray may be immediately blocked, or it may travel until hitting another part of the obstacle boundary.
2) Place one vertex in the interior of every trapezoid. It doesn't really matter where; for simplicity, pick the centroid.
3) Place one vertex in every vertical segment. The resulting vertices are shown in Figure 5(c).
4) Connect each segment vertex to the two vertices that are in the interior of the neighboring trapezoids. Each connection forms an edge in the graph and corresponds to a straight-line path.

The result is a road map that appears to capture the structure of $\mathcal{C}_{\text{free}}$. How would you implement these steps? For the first step, we could iterate over each vertex and

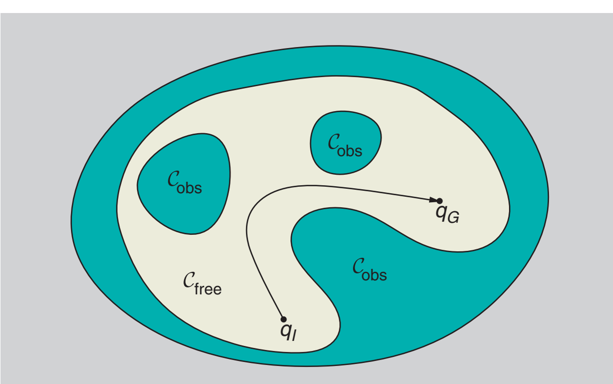


Figure 4. In the C-space, the problem looks simple: connect q_I to q_G while remaining in $\mathcal{C}_{\text{free}}$.

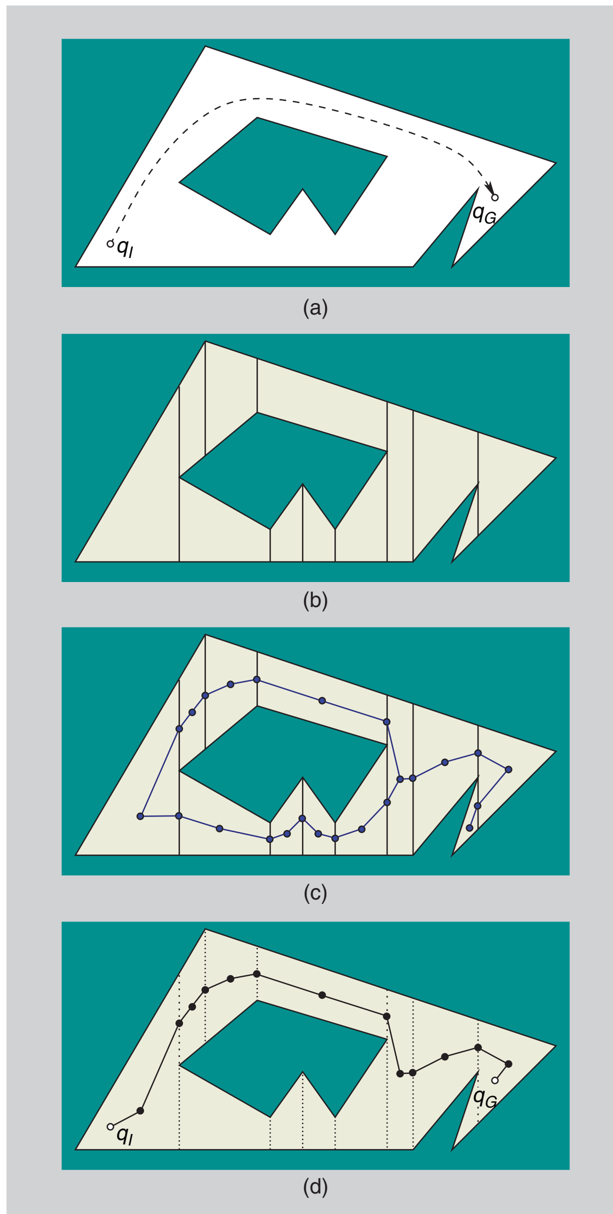


Figure 5. A combinatorial planning illustration: a) 2-D polygonal obstacle region with proposed q_I and q_G (one possible solution is shown in a dashed path); b) the trapezoidal decomposition; c) constructing a graph by placing a vertex in every vertical edge segment and every trapezoid interior; and d) connecting q_I and q_G to the graph and searching for a solution path.

determine precisely where each upward and downward ray intersects other segments. We could then easily identify the first segment hit by the vertical ray in the above and below directions. For an example as simple as Figure 5(a), this is a fine method. However, if there are n polygonal edges in total and n is large (say, $n = 20,000$), then the method is not efficient because it takes time $O(n^2)$.

By proceeding carefully, this computation can be reduced to time $O(n \lg n)$ by employing the plane sweep principle [6], which underlies many decomposition algorithms used for combinatorial planning. First, sort the polygon vertices from left to right, requiring time $O(n \ln n)$. During the algorithm execution, a list of some polygon segments is maintained and sorted from top to bottom, as they are stabbed by a vertical line. The method proceeds incrementally from vertex to vertex, traveling from left to right. At each step, the edge list is updated by simple insertions and deletions, which each take $O(\lg n)$ time using self-balancing binary search trees. If the edges incident to the vertex are both to the left, then the two edges are deleted from the list. If they are both to the right, they are inserted into the list (in order). Otherwise, the one to the left is deleted, and the one to the right is inserted. Thanks to this ordering, and we can determine in $O(\lg n)$ time the segments directly above and below the vertex, which are first stabbed by upward and downward rays. It is furthermore simple and efficient to incrementally extend the graph as each vertex is processed. For more details, see Section 6.2.2 of [9] or Section 6.1 of [6].

The road map is constructed without considering the query pair q_I and q_G. Once the investment is made, the same road map can be used for multiple query pairs. In other words, we can easily solve numerous motion-planning problems in a world that contains the same obstacle and robot. Here is a simple way to use the computed road map from Figure 5:

1) find the trapezoids that contain q_I and q_G
2) connect q_I and q_G to the vertices in their respective trapezoids
3) search the graph for a path that connects q_I to q_G.

The first step can be performed trivially in $O(n)$ time by testing whether q_I (or q_G) lies in each trapezoid; this can be shaved down to $O(\lg n)$ time by developing clever hierarchical point-location data structures [6]. The second step takes constant time, and the final step can be performed in $O(n)$ time using simple graph search algorithms such as breath first or depth first.

For the simple case of a point robot in a polygonal world, numerous alternative algorithms exist that yield comparable performance. We could, for example, decompose $\mathcal{C}_{free}$ into triangles instead of trapezoids. The general principles are that each cell should be easy to traverse (convex is ideal), the decomposition into cells should be easily computable, and the adjacencies between cells should be straightforward to determine. Based on these properties, a useful road map is obtained.

Road maps need not be obtained by cell decompositions. For example, a shortest path road map yields distance-optimal paths and is constructed by connecting certain pairs of vertices that can see each other, and each has an interior angle greater than π. A maximum clearance road map can also be computed efficiently. In general, a road map is expected to have two properties to be useful for planning:

1) *Accessibility*: It is simple to reach a point on the road map from any $q \in \mathcal{C}_{free}$ while trivially avoiding collisions.
2) *Connectivity preserving*: For any pair q_1, q_2 of points that is connected to the road map, a path exists between them in the road map if and only if there was a path between q_1 and q_2. In other words, if q_2 is generally reachable from q_1, then traveling between them via the road map must also be possible.

It seems up to this point that combinatorial planning solutions have beautiful properties. Most importantly, they construct a discrete representation of the problem that exactly captures the solution. In other words, there are no approximation or sampling errors. These methods are called *complete*, meaning that, for any input problem, they correctly determine in finite time whether or not a solution exists.

Here comes the trouble. Most motion-planning problems involve robots that are not modeled as points and they can rotate in addition to translating. How many of these nice combinatorial planning ideas extend? First, consider the case of a polygonal translation-only robot. If the robot $\mathcal{A}$ and obstacle $\mathcal{O}$ are convex polygons, then $\mathcal{C}_{obs}$ is a polygon in which every edge corresponds to a point-to-edge contact between $\mathcal{A}$ and $\mathcal{O}$. See Figures 6 and 7. Can you see how to achieve this by reassembling the edges of $\mathcal{A}$ and $\mathcal{O}$ into $\mathcal{C}_{obs}$, with the edges appearing in an ordering with the edge normals? Once this conversion is made, a trapezoidal decomposition approach is easily applied. If $\mathcal{A}$ and $\mathcal{O}$ are nonconvex, then they need to be first

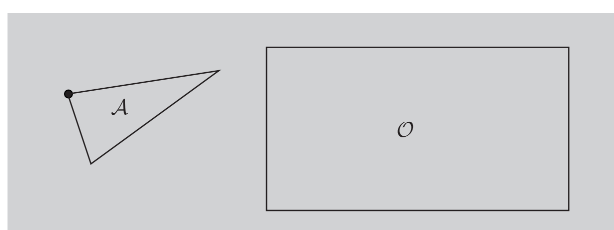


Figure 6. A triangular robot and a rectangular obstacle.

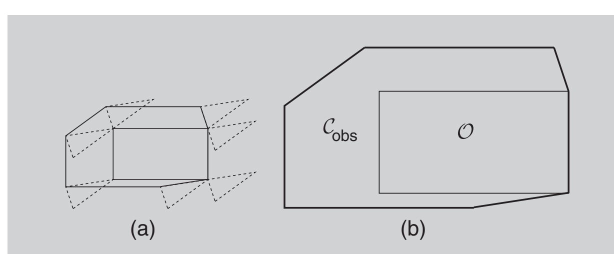


Figure 7. (a) Slide the robot around the obstacle while keeping

```
RRT(q_0)
 1  G.init(q_0);
 2  repeat
 3      q_rand → RANDOM_CONFIG(C)
 4      q_near ← NEAREST(G,q_rand);
 5      G.add_edge(q_near,q_rand);
```

Figure 8. A simple outline of the RRT algorithm.

decomposed into convex pieces to construct the convex pieces of $\mathcal{C}_{obs}$. A trapezoidal decomposition algorithm could even be used for the convex decomposition of $\mathcal{A}$ and $\mathcal{O}$.

Now introduce rotation. For the translation-only case, $\mathcal{C}_{free}$ has a piecewise linear boundary because the translation is a linear transformation. Unfortunately, the rotation is nonlinear and commonly represented using trigonometric functions. Various ways to reparameterize rotation matrices lead to improvements; however, nonlinearity is unavoidable. For computation, polynomial parametrizations are preferred. The previous piecewise-linear representations are then replaced with semialgebraic representations, meaning that each facet of $\mathcal{A}$, $\mathcal{O}$, and $\mathcal{C}_{obs}$ is represented as the roots of implicit polynomials. Constructing $\mathcal{C}_{obs}$ in terms of polynomial roots is straightforward, but a combinatorial explosion occurs that produces far too many facets for practice (the example in Figure 6 already produces more than 70). For 3-D problems, it becomes considerably worse. The next difficulty is to perform cell decomposition. The first motion-planning method to accomplish this is the cylindrical decomposition method of Schwartz and Sharir [13], which produces a number of cells that is doubly exponential in the dimension of $\mathcal{C}$. More efficient cell decomposition methods exist, and there is Canny's algorithm [3], which directly produces a road map through $\mathcal{C}_{free}$ in a singly exponential time without a prior decomposition. These methods provide solutions to the general path-planning problem; however, they are even rarely implemented due to numerical issues and inefficiency from the combinatorial explosion.

45 Iterations 2,345 Iterations

Figure 9. In the early iterations, the RRT quickly reaches the unexplored parts. However, the RRT is dense in the limit (with probability one), which means that it gets arbitrarily close to any point in the space.

Sampling-Based Planning

Sampling-based approaches are by far the most common choice for industrial-grade problems, because $\mathcal{C}_{obs}$ is composed of an unwieldy number of facets. They abandon the idea of explicitly characterizing $\mathcal{C}_{free}$ and $\mathcal{C}_{obs}$ and essentially leave the planning algorithm in the dark when exploring $\mathcal{C}_{free}$. The only light is provided by a collision-detection algorithm, which is a black box that probes $\mathcal{C}$ to determine whether some configuration (or a small ball around it) lies in $\mathcal{C}_{free}$. These algorithms often work by hierarchically representing $\mathcal{A}$ and $\mathcal{O}$ and attempting to quickly determine collision at a course resolution [11]. Many collision detection methods are incremental, which means that they can yield extremely fast performance by saving information from a previous execution on a nearby configuration.

Planning algorithms then work by incrementally probing and searching $\mathcal{C}_{free}$ for a path, gradually revealing more and more of it with the collision detector. In this way, motion planning feels like using a robot with a weak sensor to explore an unknown environment. This might seem odd since $\mathcal{O}$ and $\mathcal{A}$ are given; however, the environment being explored is $\mathcal{C}_{free}$ (or equivalently, $\mathcal{C}_{obs}$), which is high dimensional and prohibitive to explicitly represent. Sampling-based approaches attempt to find a solution quickly while cheating their way out of building a full map of $\mathcal{C}_{free}$. Don't compute more than you have to.

To get a feeling for sampling-based planning issues, we first introduce a frequently used method based on rapidly exploring random trees (RRTs). Figures 8 and 9 show the algorithm and its result. The idea is to aggressively probe and explore the C-space by expanding incrementally from an initial configuration q_0. The explored territory is marked by a tree rooted at q_0. Each iteration extends the tree by adding a leaf vertex and edge that connects it to the rest of the tree. Each edge is a collision-free path between two configurations. The RRT algorithm picks a point q_{rand} at random in $\mathcal{C}$ (not $\mathcal{C}_{free}$) and then tries to connect the tree to it by extending the nearest point in the tree. This biases the tree toward aggressively reaching unexplored parts of $\mathcal{C}$, but eventually settling on uniform coverage.

Some implementation details are needed to clarify Figure 8. Step 1 initializes G to contain a single vertex, corresponding to q_0 and no edges. In Step 3, a random configuration generator is used to obtain $q_{rand} \in \mathcal{C}$. A random translation could be selected uniformly from a bounded region (often an axis-aligned rectangle). A random 2-D rotation is easily obtained by randomly selecting some $\theta \in [0, 2\pi)$. It turns out that selecting a uniformly random 3-D rotation is technically more challenging. Here is an amazingly simple method. Choose three points $u_1, u_2, u_3 \in [0, 1]$ uniformly at random and then let [14]:

$$a = \sqrt{1-u_1}\sin 2\pi u_2 \qquad b = \sqrt{1-u_1}\cos 2\pi u_2$$
$$c = \sqrt{u_1}\sin 2\pi u_3 \qquad d = \sqrt{u_1}\cos 2\pi u_3 \qquad (4)$$

in the rotation matrix (2).

What does uniform random really mean for $\mathcal{C}$? Recall from the "Problem Formulation" section that the set of transformations could be expressed in numerous ways, meaning that the notion of uniform randomness appears to be arbitrary. There is, however, a well-defined notion of uniformity based on Haar measure, which is beyond this tutorial; see Section 5.2 of [9]. Intuitively, if we rotate the coordinate frame on which the rotations are defined, then the uniformity should be preserved. The methods for rotation above, including (4), achieve this.

Step 4 finds q_{near}, the closest point in G to q_{rand} (see Figure 10). What does it mean to be closest? This again depends precisely on how $\mathcal{C}$ is represented and implies that a distance function has been defined. The distance function $\rho : \mathcal{C} \times \mathcal{C} :\rightarrow [0, \infty)$ is formally called *metric* and usually satisfies the following axioms for all $p, q, r \in \mathcal{C}$: 1) $\rho(p, q) \geq 0$, 2) $\rho(p, q) = 0$ if and only if $p = q$, 3) $\rho(p, q) = \rho(q, p)$, and 4) $\rho(p, q) + \rho(q, r) \geq \rho(p, r)$. In virtually all sampling-based planning algorithms, performance depends on the choice of the metric. It is sometimes difficult to set the relative weights between rotational distances and translational distances (see Figure 11).

Now that the closest has been established, which points in G are checked for being the nearest to q_{rand}? The simplest is check the vertices and report the nearest one. But the closest point among all those explored could lie along an edge. Rather than incurring an expensive computational cost, a common tradeoff is to check some intermediate points at regular intervals along an edge (Figure 12). This introduces an unfortunate parameter to tune but often simplifies implementations (it is also reasonable to avoid all of this and just use the vertices).

Finally, Step 5 extends the tree. If $\mathcal{C}_{\text{obs}}$ were empty, then an edge can be made from q_{near} to q_{rand}. If q_{near} is a vertex in G, then the endpoints of the new edge are q_{near} and q_{rand}. If q_{near} is a point along the interior of an edge, then that edge must first be split, with q_{near} introduced as an intermediate vertex. Since $\mathcal{C}_{\text{obs}}$ is usually not empty, there are two issues: 1) A collision-detection algorithm makes sure that we can travel from q_{near} toward q_{rand} while staying in $\mathcal{C}_{\text{free}}$, and 2) we might not be able to reach q_{rand} without hitting $\mathcal{C}_{\text{obs}}$. If it is not possible to reach q_{rand}, then the new vertex is instead placed at the configuration q_i that gets as close as possible, as shown in Figure 13. (If no progress is possible, then no new edge and vertex are created.)

The RRT algorithm presented in Figure 8 aggressively explores $\mathcal{C}_{\text{free}}$; however, if the tree is grown from q_I, there is no consideration of q_G. Now consider ways to solve the basic path-planning problem using RRTs.

Here is a simple adaptation. Start the RRT with $q_0 = q_I$, and at every 100th iteration, force $q_{\text{rand}} := q_G$ instead of choosing a random configuration. If q_G is reached, then a path has been found from q_I to q_G, which solves the problem. This induces a gentle bias toward the goal. At one extreme, we could pick q_G every time, making a beeline for q_G. This would fail miserably when an obstacle is reached. Figure 14(a) shows an example in which this would occur. Aggressively attempting to reach q_G by setting $q_{\text{rand}} := q_G$ in every other iteration would still work, but might waste too much effort running into $\mathcal{C}_{\text{obs}}$ instead of exploring. Therefore, a light bias, such as every 100th iteration is recommended.

For many problems, though, such a simple strategy is not enough. Figure 14(b) shows a kind of bug trap from which it is difficult to escape. Because of the existence of

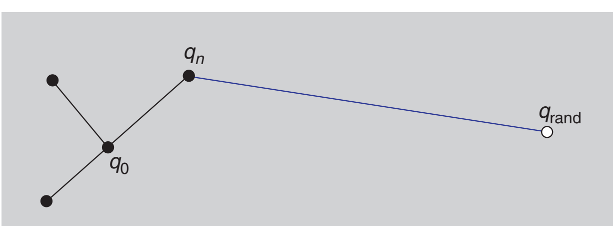


Figure 10. A new edge is added that connects from the random sample q_{rand} to the nearest point in S, which is the vertex q_n.

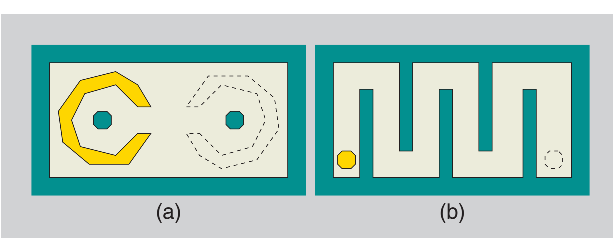


Figure 11. Rotation versus translation domination: (a) The task is to move the C shape to the right. Rotation dominates. Performance should improve if rotation is weighted heavily in the metric. (b) In this case, the translation dominates and should therefore be weighted more heavily if this fact is known in advance.

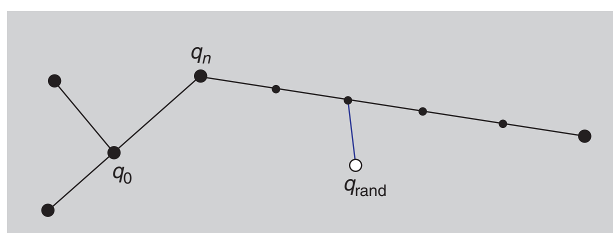


Figure 12. For ease of implementation, intermediate vertices can be inserted to avoid checking for the closest points along line segments. The tradeoff is that the number of vertices is increased dramatically.

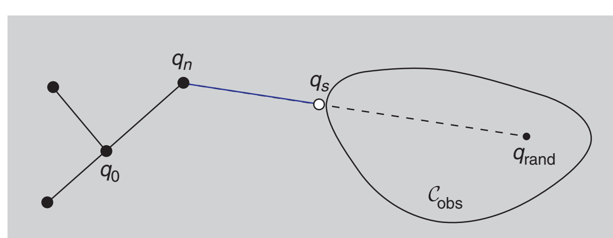


Figure 13. If there is an obstacle, the edge travels up to the obstacle boundary, as far as allowed by the collision-detection algorithm.

IEEE Robotics & Automation Magazine | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

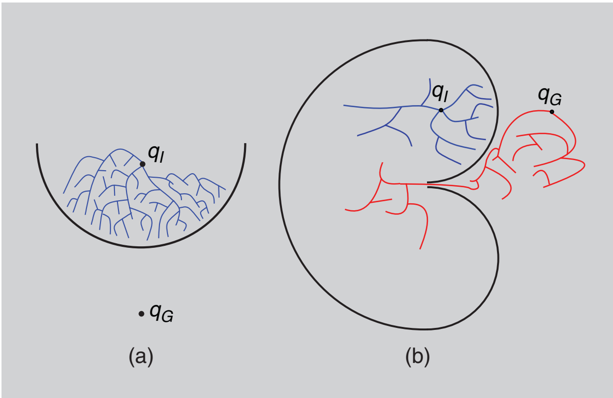


Figure 14. The C-space obstacles may contain wells that trap planners in local minima or one-way doors that resemble bug traps. (a) Filling a well. (b) A bug trap.

such situations, which commonly occur in practice, a bidirectional search is more effective and popular. The algorithm grows two RRTs: 1) G_I rooted at q_I and 2) G_G rooted at q_G. Instead of always extending the trees using random configurations, half of the time is spent trying to extend each tree toward the newest vertex of the other tree. The following four iterations are repeated:

1) generate q_{rand} and use it to extend G_I, obtaining a new leaf vertex q_{new}
2) force $q_{rand} := q_{new}$ and use it to extend G_G
3) generate a new q_{rand} and use it to extend G_G, obtaining a new leaf vertex q_{new}
4) force $q_{rand} := q_{new}$ and use it to extend G_I.

Steps 1 and 3 are identical to the execution in Figure 8, but for G_I and G_G, respectively. Steps 2 and 4 trick the RRT by using the most recent vertex from the other tree as a replacement for q_{rand}. If either of these two steps ever succeed in connecting the trees to each other, then the problem is solved. This method is quite effective for most practical problems, as aggressive exploration from q_I and q_G is balanced with trying to connect the trees to solve the problem.

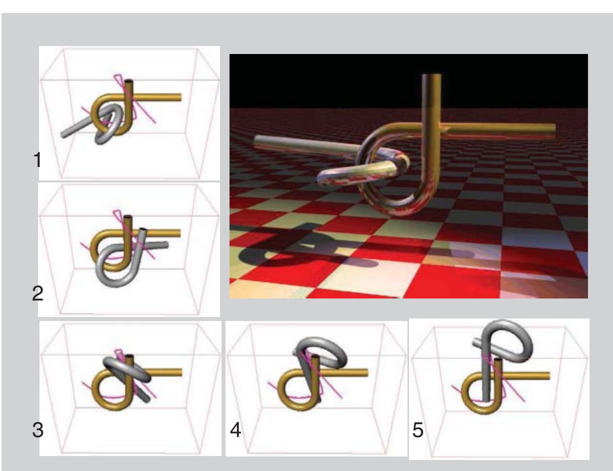


Figure 15. The bidirectional RRT solves the Alpha 1.0 puzzle in a few minutes.

An example that was solved in 2002 by the bidirectional RRT is the famous Alpha 1.0 puzzle introduced by Nancy Amato and Boris Yamrom. The task is to pull apart the twisted nails, leading to an extremely narrow corridor in $\mathcal{C}_{free}$ through which the solution path must travel. The solution is illustrated in Figure 15. Most problems are not this challenging, and solutions are often found in a fraction of a second. Nevertheless, there are limitations to the method as well as any sampling-based method. It is not hard to construct pathological examples that cause the algorithm to converge too slowly. In some cases, problem-specific heuristics can then be developed to recover performance.

The RRT-based methods fall into a larger family of methods called *incremental sampling and searching*, in which a graph is incrementally constructed inside of $\mathcal{C}_{free}$. Each method has a vertex selection method, which determines where to expand next from among vertices in the graph. After that, a local planning method constructs an edge from the selected vertex, thereby extending the tree. In the case of an RRT, the vertex selection method picks the vertex closest to q_{rand}. The local planning method attempts to connect the vertex to q_{rand}. As an example of an alternative incremental sampling and searching method, the expansive space planner (ESP) [7] selects a vertex with probability that is inversely proportional to the number of other vertices within a ball of predetermined size. The local planning method then connects to a random configuration within the ball, but only with a probability that is inversely proportional to the number of vertices that lie within a ball centered on the random configuration. Another example that falls into this family is the randomized potential field planner [2], which implements gradient descent in $\mathcal{C}_{free}$ and uses random walks to escape local minima.

A common nuisance with sampling-based planning methods is that the produced paths are jagged as they traverse $\mathcal{C}_{free}$. This makes the solution animation jumpy; Making the robots to follow such awkward paths is a comically bad idea. Therefore, path smoothing is usually performed to clean up solution paths. Fortunately, it is straightforward to produce a cleaner path once a jagged solution is given. A simple method is to iteratively pick a pair of points at random along the path and attempt to replace the path portion between them with a straight line in $\mathcal{C}_{free}$. If this survives the collision-detection verification step, then use the linear segment and discard the original part portion. After several dozen iterations, the path is usually much improved.

The discussion so far has focused only on single-query algorithms, meaning that only one q_I, q_G pair will be given so that there are no advantages of extensive precomputation. Recall from the "Combinatorial Planning" section that planning problems can be quickly solved once a nice road map has been computed that offers the accessibility and connectivity-preserving properties. This motivates a multiple-query approach to sampling-based planning known as a *probabilistic road map* [8]. In this case, a bunch (e.g., 1,000) of random

Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

configurations are chosen upfront and declared to be road map vertices. Road map edges are formed by attempting to connect each configuration to all vertices within some specified radius (Figure 16). If a road map can be constructed that satisfies accessibility and connectivity preservation with high probability, then it can be used to efficiently search for solutions to multiple initial-goal query pairs. One difficulty is that the road map may have as many edges and vertices as a high-dimensional grid [10], which provides motivation for pruning strategies that attempt to keep the good road map properties while reducing its size substantially. See, for example, the visibility road map variant [15].

To conclude, we should emphasize that a tradeoff has been made by going to sampling-based methods. Recall from the "Combinatorial Planning" section that combinatorial planning leads to complete algorithms: They always find a solution if it exists; otherwise, they report failure. Since sampling-based methods solve problems without fully characterizing $\mathcal{C}_{obs}$, completeness is reduced to weaker forms. The goal is to ensure that the sampling eventually covers all of $\mathcal{C}$. This can be expressed in terms of dispersion, which is the radius of the largest empty (unsampled) ball in $\mathcal{C}$. Sampling-based approaches usually achieve resolution completeness, meaning that they will find a solution if one exists, but may run forever if one does not, or probabilistic completeness, meaning that the probability tends to one that a solution is found if one exists (otherwise, it may still run forever). For example, the RRT approaches described above lead to probabilistic completeness, partly because the dispersion is reduced to zero with probability one. Resolution completeness can be obtained by replacing the random configuration generator by a deterministic point sequence that leads to zero dispersion in $\mathcal{C}$ in the limit (for example, consider a multiresolution grid that refines forever).

The best way to learn more about sampling-based motion planning is to experiment with the implementations. You could download and install a free library, such as the Open Motion Planning Library from Rice University, the Motion Strategy Library from the University of Illinois, or the Motion Planning Kit from Stanford. If you instead want to start from the basics, then at least downloading a collision-detection package, such as PQP from the University of North Carolina, is recommended.

Direct Extensions

Now that the core motion-planning ideas have been explained for the case of rigid 2-D or 3-D robots among fixed obstacles, several straightforward extensions can be covered for which the planning methods are virtually the same.

The formulation given in the "Problem Formulation" section allowed only one moving rigid body. This limited the C-space to having no more than dimension three for $\mathcal{W} = \mathbb{R}^2$ and six for $\mathcal{W} = \mathbb{R}^3$. If we allow multiple moving bodies, then there is no limit on the degrees of freedom, and hence, the dimension of $\mathcal{C}$. Consider, for example, Figure 17, in which a bunch of rectangles need to be rearranged by translation only. Each contributes 2-D to $\mathcal{C}$. Interestingly, this problem is already NP-hard (and PSPACE-hard) if there is no maximum limit on the number of rectangles. (If the dimension of $\mathcal{C}$ is bounded in advance, then the path-planning problem is solvable in time polynomial in the representation of the robot and world obstacles.)

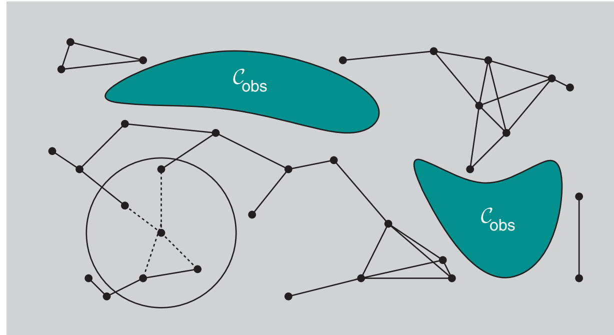


Figure 16. The probabilistic road map method attempt to achieve road map accessibility and connectivity preservation via random sampling and connecting to nearby samples.

Planning a collision-free path for multiple rigid bodies is no different conceptually to planning for a single body, once we think in terms of $\mathcal{C}$ and $\mathcal{C}_{free}$. The configuration vector $q \in \mathcal{C}$ includes coordinates to place each body. For example, for two translation-only rectangles, $q = (x_1, y_1, x_2, y_2)$ represents their position and $\mathcal{C} = \mathbb{R}^4$. The initial q_I and goal q_G configurations now express the placement of every body. Suppose there are n bodies $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$, with configuration parameters $q_1, \ldots, q_n$. If $\mathcal{A}_i$ is transformed into configuration q_i, it occupies $\mathcal{A}_i(q_i) \subset \mathcal{W}$ in the world. Let $q = (q_1, \ldots, q_n)$ represent the simultaneous configuration of all bodies. A configuration is collision free, $q \in \mathcal{C}_{free}$, if and only if $\mathcal{A}_i(q_i) \cap \mathcal{O} = \emptyset$ for every i from 1 to n, and $\mathcal{A}_i(q_i) \cap \mathcal{A}_j(q_j) = \emptyset$ for every $i \neq j$. In other words, for $q \in \mathcal{C}_{free}$, there must be no body–obstacle collisions and no body–body collisions.

Once $\mathcal{C}$, q_I, q_G, and $\mathcal{C}_{free}$ are defined in this way, the methods given in "Combinatorial Planning" and

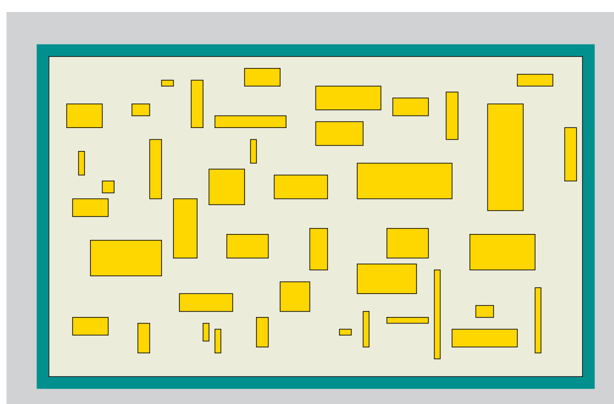


Figure 17. Consider rearranging many rectangles, with no rotations, inside of a rectangular box in $\mathbb{R}^2$. Without a limit on the number of rectangles, the problem is NP-hard.

"Sampling-Based Planning" sections directly apply. The only difficulty is that the dimension of $\mathcal{C}$ is large, which limits the applicability of combinatorial methods and some sampling-based methods. This has motivated the development of various decoupled approaches, which avoid considering all bodies at once. For example, paths may be planned for each body individually, and then their motions along the paths can be set correctly so that collisions are avoided. Such methods are not complete but are practical in many settings. Alternatively, dimensionality-reduction techniques, such as those based on the Johnson-Lindenstrauss Lemma, may hold promise for adapting sampling-based planning methods to directly account for all bodies simultaneously.

Figure 18. The classic Puma 560 arm is a chain of three rotatable bodies (excluding the end effector) attached to a rigid base. This yields a three-dimensional C-space, which is handled by the standard planning algorithms. (Photo courtesy of the Technical University of Berlin.)

If bodies are allowed to contact each other, several other motion-planning variants are obtained. Two will be considered here: 1) articulated bodies and 2) manipulation. For articulated bodies, they are attached together by joints that enable some freedom of motion between them, as shown in Figures 18 and 19. The attachment of bodies removes some of their collective degrees of freedom. Configuration coordinates express how each body is situated with respect to bodies to which it is connected. Expressions for transforming such bodies are just standard robot kinematics covered in numerous textbooks [5], [16]. Somewhat different from standard kinematics, we are once again interested in the set of all possible transformations, resulting in the C-space. Once this has been defined, a manifold C-space $\mathcal{C}$ is usually obtained, on which q_I, q_G, and $\mathcal{C}_{free}$ are straightforward to define. Here, $\mathcal{C}_{free}$ includes some configurations in which there are body–body collisions, but only if these they are attached by a joint. Once defined, the methods of "Combinatorial Planning" and "Sampling-Based Planning" sections once again apply, with the usual warning about the dimension of $\mathcal{C}$.

A more serious complication is when a collection of articulated bodies forms a loop, as shown in Figure 20. The result is called a *closed kinematic chain*, which occurs in parallel robots and if multiple robots contact the same body for manipulation. In most cases, it is difficult to explicitly characterize the set of configurations that satisfy the loop-closure constraint. This makes it difficult to even parameterize paths through $\mathcal{C}$. Sampling-based planning approaches have nevertheless been developed to step through this difficult space by ensuring that loop closure is maintained while incrementally searching for a solution path.

Manipulation problems more generally require robots to determine which bodies to grasp and how to carry them to solve a problem. For example, the task might be to use a manipulator arm to stack several boxes. The degrees of freedom of boxes in addition to the robot are all included when defining $\mathcal{C}$. The task is expressed by specifying a configuration in which the boxes are stacked. This problem conceptually appears more challenging. Standard algorithms are often adapted to solve it by forming a hybrid C-space that includes discrete variables in addition to configuration variables. The discrete variables record modes of interaction. For example, there is a transit mode, when the manipulator is not carrying a body, and a transfer mode, when it carries a body. Heuristics are then used to determine when modes should be switched, in addition to solving the planning problem that arises in each mode.

Another variant of the basic path-planning problem is to allow the obstacles to move. Let $T = [0, t_f]$ be an interval of time, in which t_f is some final time. In this case, a snapshot of the world can be imagined at every time $t \in T$. The obstacle region $\mathcal{O}$ becomes $\mathcal{O}(t)$. Now consider computing a collision-free path from time $t = 0$ to time $t = t_f$. This is

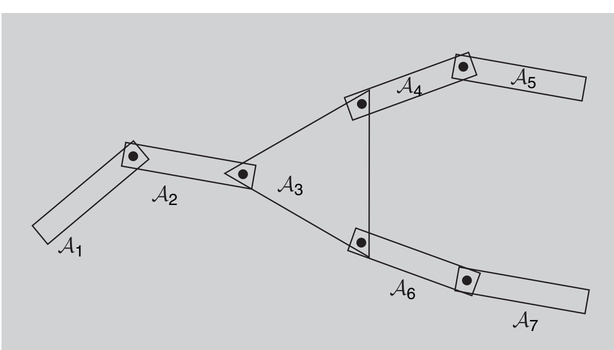


Figure 19. Seven links are attached via rotatable joints. If each is allowed a full range of motion from 0 to 2π, then $\mathcal{C}$ is a seven-dimensional torus.

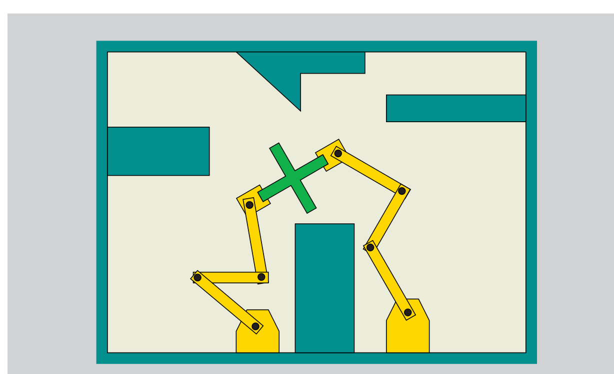


Figure 20. Two or more arms manipulating the same object causes a closed kinematic chain.

IEEE Robotics & Automation Magazine
Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

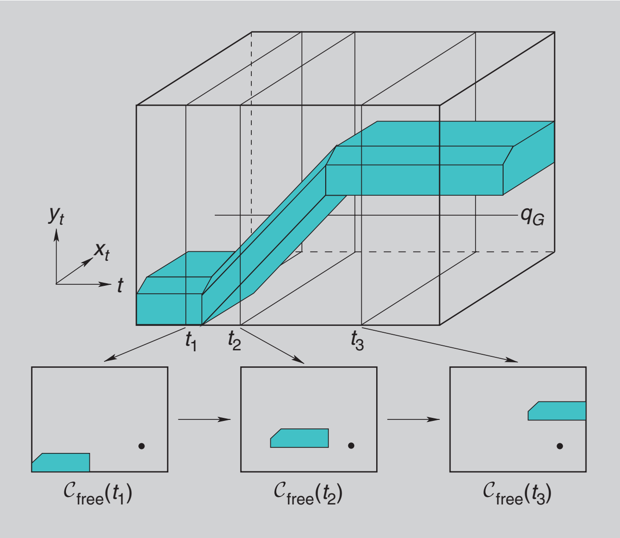


Figure 21. A time-varying example with piecewise-linear obstacle motion. Planning through the state-time space occurs.

conceptually straightforward if we construct the configuration-time space, $Z = \mathcal{C} \times T$. Figure 21 shows an example of how this appears. To solve the problem, the path-problem algorithms work in the usual way with one exception: The path must always make forward progress through time. The combinatorial road map methods and incremental sampling and searching methods can be adapted without much difficulty to enforce this. It becomes considerably more challenging, however, if the robot has a maximum speed bound. This yields a constraint on the path slope through Z, which is more difficult to enforce. Finally, it is even more difficult and practical, when there is uncertainty in predicting the future motions of the obstacles. This falls under the topic of uncertainty, which is covered in the next tutorial part.

Conclusions

After reading this, you should hopefully have extracted the following main points. Motion planning lives in the C-space, which is the set of all transformations. Combinatorial planning solves simpler problems in a clean, elegant way, but the running time is too high for industrial-grade problems. Sampling-based planning provides practical solutions for real-world problems but offers weaker guarantees. Performance degrades for problems in which narrow doorways in $\mathcal{C}_{free}$ are hard to find. Several extensions to the standard path-planning problem expand the C-space definition and require only minor adaptations to the usual approaches. The key issue is that the C-space dimension increases, which generally raises computational complexity.

So we have seen powerful methods that generate a collision-free path automatically. Not bad. This is useful in many settings, extending well beyond robotics. But what if a robot is not able to follow the path due to differential constraints arising from kinematics and dynamics? What if we cannot predict precisely where the robot will go? What if the obstacle locations are uncertain and possibly changing? These concerns, with which every roboticist is familiar, motivate the topics in the second part of this tutorial.

References

[1] V. I. Arnold, *Mathematical Methods of Classical Mechanics*, 2nd ed. Berlin: Springer-Verlag, 1989.

[2] J. Barraquand, B. Langlois, and J. C. Latombe, "Numerical potential field techniques for robot path planning," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 2, pp. 224–241, 1992.

[3] J. F. Canny, *The Complexity of Robot Motion Planning*. Cambridge, MA: MIT Press, 1988.

[4] H. Choset, K. M. Lynch, S. Hutchinson, G. Kantor, W. Burgard, L. E. Kavraki, and S. Thrun, *Principles of Robot Motion: Theory, Algorithms, and Implementations*. Cambridge, MA: MIT Press, 2005.

[5] J. J. Craig, *Introduction to Robotics*. Reading, MA: Addison-Wesley, 1989.

[6] M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf, *Computational Geometry: Algorithms and Applications*, 2nd ed. Berlin: Springer-Verlag, 2000.

[7] D. Hsu, J.-C. Latombe, and R. Motwani, "Path planning in expansive configuration spaces," *Int. J. Comput. Geometry Applicat.*, vol. 4, pp. 495–512, 1999.

[8] L. E. Kavraki, P. Svestka, J.-C. Latombe, and M. H. Overmars, "Probabilistic roadmaps for path planning in high-dimensional configuration spaces," *IEEE Trans. Robot. Automat.*, vol. 12, no. 4, pp. 566–580, June 1996.

[9] S. M. LaValle. (2006). *Planning Algorithms*, Cambridge, U.K., Cambridge University Press [Online]. Available: http://planning.cs.uiuc.edu/

[10] S. M. LaValle, M. S. Branicky, and S. R. Lindemann, "On the relationship between classical grid search and probabilistic roadmaps," *Int. J. Robot. Res.*, vol. 23, no. 7/8, pp. 673–692, July/Aug. 2004.

[11] M. C. Lin and D. Manocha, "Collision and proximity queries," *Handbook of Discrete and Computational Geometry*, 2nd ed., J. E. Goodman and J. O'Rourke, Eds. New York, Chapman and Hall, 2004, pp. 787–807.

[12] T. Lozano-Pérez, "Spatial planning: A configuration space approach," *IEEE Trans. Comput.*, vol. C-32, no. 2, pp. 108–120, 1983.

[13] J. T. Schwartz and M. Sharir, "On the piano movers' problem: III. Coordinating the motion of several independent bodies," *Int. J. Robot. Res.*, vol. 2, no. 3, pp. 97–140, 1983.

[14] K. Shoemake, "Uniform random rotations," in *Graphics Gems III*, New York: Academic, 1992, pp. 124–132.

[15] T. Siméon, J.-P. Laumond, and C. Nissoux, "Visibility based probabilistic roadmaps for motion planning," *Adv. Robot. J.*, vol. 14, no. 6, 2000.

[16] M. W. Spong, S. Hutchinson, and M. Vidyasagar, *Robot Modeling and Control*. New York: Wiley, 2005.

[17] S. Udupa, "Collision detection and avoidance in computer controlled manipulators," Ph.D. dissertation, Dept. Elect. Eng., California Inst. Technol. 1977.

Biography

Steven M. LaValle Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL lavalle@uiuc.edu.

Impact of SAC

By Tamas Haidegger

The first year of our term has passed, and the Student Activities Committee (SAC) has been trying to make an impact: boosting your conference experience at the IEEE International Conference on Robotics and Automation (ICRA) and the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) and providing alternative ways to develop your professional career to network and improve your professional skills. You can read a short report on the activities of the IROS below from Michel Franken (see "Sneak Peek into Our Upcoming ICRA 2011 Events"). Michel was one of my cochairs in 2010, and he did an amazing job, pulling together SAC activities. I am happy to announce here that for his service he has been awarded the IEEE Robotics Automation Society (RAS) Outstanding Student Volunteer Award. Congratulations!

Meanwhile, we are eagerly looking forward to the next year, focusing on enlarging the scope and attendance of our programs and reaching out to a larger number of students.

Back in December, we started our preparation for ICRA 2011 to make it an even larger event. We hope to see most of you at the conference and at our programs. You can find more details about the tentative program in "Sneak Peek into Our Upcoming ICRA 2011 Events," but please follow the official Web site for updates.

In addition, you can read a short notice from Alejandro, my other hardworking cochair, on academic career development (see "Academic Career Advice from Your Future Self"), an inspiring edited interview with a couple of the recent best student paper award winners, and finally, the first article in a new series from SAC on the brand new Student Reviewing Program (see "The Reviewing Process: An Introduction for New Reviewers"), coordinated by Ludo. Do not miss it.

Finally, you are most welcome to join our team. For more details, check out our Web site: http://wiki.ieee-ras.org/mab/sac or e-mail me at ras_sac@ieee.org.

Digital Object Identifier 10.1109/MRA.2011.940897
Date of publication: 14 April 2011

What's Behind the Best Papers

By Alejandro Perez and Tamas Haidegger

Interview with the Recent Recipients of Best Awards

Without a doubt, research publications are one of the major driving forces behind the scientific progress in our field. Conducting research that provides significant results and leads to materials worthy of publication and presentation is a challenging, but widely known and understood process. However, we can always learn from good examples, analyzing what makes certain papers outstanding in their research area, recognized by awards. In this article, you can read an edited interview with young professionals who were recently honored for their excellent papers. They tell us about the background on how they achieved it.

Three articles are featured:

- *Best Medical Robotics Paper, ICRA 2010*: "Superhuman Performance of Surgical Tasks by Robots Using Iterative Learning from Human-Guided Demonstrations" by Jur van den Berg et al.
- *Best Conference Paper, RSS 2010*: "Biophysically Inspired Development of a Sand-Swimming Robot" by Daniel I. Goldman et al.
- *Best Student Paper, RSS 2010*: "Passive Torque Regulation in an Underactuated Flapping Wing Robotic Insect" by Pratheev Sreetharan et al.

Digital Object Identifier 10.1109/MRA.2011.940898
Date of publication: 14 April 2011

How did the research group work together? What were the main contributions of each author?

van den Berg: We worked on this project with a large group. The runup to the final result was a long process, which started with getting back into operation a 13-year-old laparoscopic robotic platform. Getting it to work can mainly be attributed to Andrew Wan, Humphrey Hu, and Xiao-Yu Fu.

IEEE Robotics & Automation MAGAZINE | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

Sneak Peek into Our Upcoming ICRA 2011 Events

Fostering Interaction Between Roboticists and Students

Student/Chair Mentorship Program

At ICRA 2011, students can get involved in a fun way again. Get the behind-the-scenes experience of a conference. This program will give you a chance to interact with key researchers in your field. Students will be paired with a session chair, where you will learn how to run a session. Students who are interested should first find out who they would like to work with by reading the session guide and selecting a chair that they would like to be their mentor. Please e-mail the name of the session, the date, and your relevant contact information to fibrs@ieee.org. Keep in mind that the people are occasionally very busy, so you may also wish to provide alternative mentor names. Pick topics you are interested in rather than focusing on famous names.

Student Photo Contest

You are welcome to submit your photographs taken during any professional event. This is a seasonal amateur photography competition, which is open to undergraduate and graduate student members. The judging of the submitted photographs will be made by the IEEE RAS SAC, involving independent judges. Three winners will be announced and awarded after the conference. The submitted photos will be used for archival purposes.

Student Reporters Program

You are given a chance to become famous with your writing skills: we are calling for entertaining, yet professionally relevant reports on different workshops and conferences. Specifically for ICRA reporters, you can register yourself to a session via SAC Web site (http://wiki.ieee-ras.org/mab/sac). The best reports will be awarded and published on the RAS Web site and/or in an upcoming issue of *IEEE Robotics & Automation Magazine*. Contact RAS_student_reporters@ieee.org for more details.

Explore Shanghai Beyond Pudong

SAC is organizing tours in the free timeslots at ICRA to explore Shanghai's day and night. We will have local student guides to show us the most interesting places, traditional food, and local drinks.

Tamas Haidegger

Once operational, the focused research on learning knot ties and other procedures through human demonstrations was mainly carried out by Jur van den Berg, Stephen Miller, and Daniel Duckworth. Jur, as a postdoctoral researcher, led the project on a daily basis. He wrote the paper and the algorithms for learning trajectories from demonstrations. Stephen and Daniel, undergraduate students, worked in the laboratory on implementing these algorithms to work on the robots. Humphrey Hu was never far away to resolve hardware issues with the robots in case they appeared. Ken Goldberg and Pieter Abbeel supervised the project on a higher level. They provided global directions and feedback on the progress.

Goldman: The research group was composed of two of my graduate students, Ryan Maladen (a bioengineering Ph.D. student) and Yang Ding (a physics Ph.D. student), an undergraduate student (Adam Kamor), and a collaborator of mine, Dr. Paul Umbanhowar. Ryan and Paul developed the robot, while Ryan, Yang, and Adam developed the experimentally validated numerical

SAC Report from IROS 2010

The IROS was held in Taipei, Taiwan, 18–22 October. The conference itself was quite interesting, with even more presentations and exhibitors than usual; the program featured a lot of technical sessions piled with innovative research and three excellent keynote speakers, Prof. Pfeiffer, Dr. Cousins, and Prof. Sankai.

As we are responsible for the student programs, SAC tried to offer some alternatives running in between the official conference schedule. Primarily, RAS students were provided with the opportunity to get to know each other. We visited Taipei 101 together (the second tallest building the world) and were pleasantly surprised by the amazing view of the city from above when the weather all of a sudden decided to clear up. We also explored the night life of Taipei and toured the famous Shilin Night Market where all kinds of foods, mostly clothes and toys were for sale. A spectacular ending of the conference was provided with a "sing along" session for three hours in a local karaoke bar. This was great fun, and there are actually some nightingales hidden amongst us, roboticists. Some of the people who were still in Taipei on Saturday joined us on a visit to the National Palace Museum, where we saw a lot of pottery, books (that unfortunately nobody there could read), and an amazing garden. In the afternoon, we explored the downtown area (Longshan temple, snake market, Peace park) and finally enjoyed the sunset at the magnificent Chiang Kai-Shek Memorial.

A new event for SAC was the Lunch with Leaders. During this special meal, the students could have a casual discussion with well-recognized professionals, including Dr. Kosuge, Dr. Pfeiffer, Dr. Cousins, Dr. Ng-Thow-Hing, Dr. Ryu, Dr. Ferre, Dr. Niemeyer, Dr. Khatib, Dr. Corke, Dr. Du Pont, Dr. Stramigioli, and many others. Based on the feedback we got, the Lunch with Leaders will definitely be continued. Do not miss it next time at ICRA 2011.

We have the winner of the IROS student photo contest, William Morris, from City College of New York. Congratulations! You can see the winning entry at http://wiki.ieee-ras.org/mab/sac/iros2010.

We also tried to run the Fostering Interaction Between Roboticists and Students (FIBRS) Program at IROS in which the students can cochair a technical session to learn more about how the major conferences are organized and to interact with a session chair of their choice. Unfortunately, there were not many student requests, and the requests eventually could not get fulfilled. To those students who we could not help, our apologies, and we expect to do better again at the next ICRA.

This was a small overview of the events SAC organized during IROS. It is great fun to interact with fellow students outside of the technical sessions, so make sure you do not miss these events the next time.

Hope to see most of you in Shanghai.

Michel Franken

PS: Should you have any feedback on our activities, do not hesitate to contact us at ras_sac@ieee.org.

simulation of the robot and granular medium. I supervised and guided the project.

Sreetharan: Our smart composite microstructure fabrication techniques enable many of the interesting millimeter-scale robotic structures produced by our research group. Progress in these techniques is highly collaborative, with improvements and refinements quickly advancing from individual experimentation to laboratory standard.

This article considers the problem of underactuated robotics in an atypical framework.

For this paper, Pratheev conceived of the PARITy methodology for control, designed the roll-torque balancing PARITy drivetrain and created the theoretical dynamic model. He also leveraged the group's existing fabrication techniques to build the experimental structures, and he conducted the experimental trials described in the paper. Prof. Robert Wood assisted with helpful discussions, material support, and with mechanical assembly.

What do you think made the paper strong and ultimately worthy of the award?

van den Berg: I think the answer to this question is a combination of factors. First, we studied a problem of high practical relevance, given the enormous growth of robotic surgery platforms over the past couple of years. Second, the algorithms at the basis of our approach to learn optimized trajectories from demonstrations and speed them up are elegant and based on strong theoretical foundations. Third, we made it work, and showed the results that are promising for future developments in this direction. In short, our paper bridged the strong theory with relevant practice, and made it work on real robots.

Goldman: The paper builds upon our biological studies (also led by Ryan Maladen) of the sand-swimming of the sandfish lizard, results reported in *Science* ("Undulatory Swimming in Sand: Subsurface Locomotion of the Sandfish Lizard" by Ryan Maladen, Yang Ding, Chen Li, and Daniel I. Goldman, *Science*, vol. 325, p. 314, 2009). In this study, we discovered how the lizard propels itself within the sand by using an undulation of its body. The robot serves as a physical model of the organism and allows us to test hypotheses about movement patterns, for example, why does the animal always use a particular amplitude of body undulation to dive into the sand? We find that the robot swims fastest when it uses this amplitude. The other merit of this work is that we were able to develop an accurate computer model of the robot—the challenge here was to create a model of the granular medium. While such models (partial differential equations called Navier-Stokes equations) are well known in fluids like air and water, the equations at this level do not exist for granular media. Therefore, we used what is called discrete element simulation to simulate the movement of hundreds of thousands of colliding spheres in the computer and validated this simulation against the experiment (measuring drag forces in experiment and simulation). The simulation agreed quite well with robot-experimental measurements (for example, the speed of robot as we varied its wave frequency, amplitude, etc.). This provides us (and future researchers) a tool that allows accurate simulation modeling of devices that must interact with sand.

Sreetharan: This paper introduces a novel control methodology for microrobotic air vehicles that breaks from conventional wing trajectory control espoused by the related work in the field. Mechanically intelligent structures, such as the one described in this article, have the potential to greatly simplify active control systems for severely mass- and power-limited airborne robotic insects, while also providing insight into passive mechanisms potentially available to biological insects.

In a broader sense, this article considers the problem of underactuated robotics in an atypical framework. Whereas traditional underactuated robotics seeks to control the state of systems with more degrees of freedom than actuators, this article analyzes how adding the degrees of freedom can actually increase the performance of an underactuated robotic system by introducing beneficial passive dynamics.

What do you consider to be the major lesson learned while working on this project?

van den Berg: One of the main lesson is, although known by everybody, that working with hardware always presents (un)pleasant surprises during experimentation. Either the behavior of the robot is suddenly unpredictable or it breaks down for no apparent reason. Dealing with these issues makes it hard to predict how much time each step in the process takes.

Goldman: The major lesson is that physical robot models and simulation models can have predictive power for biological performance once the interaction models with the environment are established.

Sreetharan: We learned the importance of exacting and methodical design. In a first prototype, an oversight led to one of the mechanical joints exceeding its maximum force rating and buckling once the device began flapping its wings at 110 Hz. In addition to addressing these concerns about device strength, we took care to control the dynamics of individual elements of the experimental structure to tight tolerances. This allowed our classical theoretical model to accurately predict the behavior of the greatly underactuated robotic system without resorting to parameter fitting.

What was the writing process like? Does the group have any particular modus operandi that is used while redacting the material to be published?

van den Berg: The writing was the main responsibility of the first author, and drafts of the final version were ready about two weeks ahead of the deadline. This gave every member of the team the chance to review the paper at a time of their convenience and suggest changes, which were then incorporated by the first author. This cycle repeated a few times, such

Humanoids 2011

11th IEEE-RAS International Conference on Humanoid Robots
Golf Hotel, Bled, Slovenia
October 26th – 28th, 2011

Important dates

- **Submission deadline**: May 29th, 2011
- **Notification of paper acceptance**: August 6th, 2011
- **Submission of final, camera-ready papers**: September 6th, 2011

Organizing committee

General conference chair:
Aleš Ude
Jožef Stefan Institute, Slovenia

General co-chairs:
Rüdiger Dillmann, KIT, Germany
Stefan Schaal, USC, USA

Program co-chairs:
Nancy Pollard, CMU, USA
Jun Morimoto, ATR, Japan
Tamim Asfour, KIT, Germany

Tutorial/Workshop co-chairs:
Georgio Metta, IIT, Italy
Aleš Leonardis, UL, Slovenia

Awards co-chairs:
Gordon Cheng, TUM, Germany
Kazuhito Yokoi, AIST, Japan

Publicity chair:
Erhan Oztop, ATR, Japan

Financial chair:
Bojan Nemec, JSI, Slovenia

Publication chair:
Andrej Gams, JSI, Slovenia

Call for papers

The 11th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2011) will be held on October 26th-28th, 2011, in the idyllic Alpine town of Bled in Slovenia. Bled is easily accessible from the Ljubljana airport, which has excellent connections to all major European hubs. The conference is sponsored by IEEE Robotics and Automation Society.

Scope

Papers are solicited in all related areas of humanoid robotics including mechatronics, control, perception, planning, learning, neuroscience, and human-robot interaction.

Contributed papers

The contributed papers can include but are not limited to the following list of topics:

- Mechanism design and control of humanoid robots
- Software and hardware architecture, system integration
- Stability and dynamics for humanoid robots
- Whole body motion planning and control
- Learning strategies for humanoid robots
- Imitation learning
- Perception for humanoid robots
- Humanoid grasping and manipulation
- Human-humanoid interaction
- Planning and cognition on humanoid robots
- Human body and behavior modeling
- Neuro-robotics and robot-brain interfaces for humanoids
- Humanoid robot applications

Accepted papers will be presented either at a single track oral session or during poster sessions. All the accepted papers will appear in the proceedings of the conference without distinction and on IEEE Xplore.

Detailed information about the conference and paper submission will be published on the conference website http://www.humanoids2011.org/ in a timely fashion. All inquires about the conference should be sent to the conference secretariat by e-mail (humanoids@ijs.si).

Tutorial & workshop proposals

Proposals for half-day and full day workshops and tutorials must be submitted by May 29th. Proposals must include 1) Title, 2) Organizers with contact information, 3) Objectives and topics, and 4) List of prospective speakers.

http://www.humanoids2011.org/

Digital Object Identifier 10.1109/MRA.2010.940164

that the end result was carefully internally reviewed and approved by every member of the team before submission.

Goldman: The writing process was pretty smooth on this paper. The student, Ryan Maladen, did a fantastic job of producing a first draft.

Sreetharan: We follow the standard procedures of ensuring that any intellectual property is adequately protected before publication. We also believe that clear and appealing imagery is at least as important to conveying a scientific work as is clear writing; thus, much care was taken to ensure that the figures were clear, informative, appealing, and polished.

Has the group continued to work on this project? What can we expect from future publications coming out of your laboratory?

van den Berg: Yes, we are continuing research in this direction. A major shortcoming of our paper was that the robots essentially operated blindly and assumed knowledge of the state of the suture if it needed to grasp it. The main questions we are working on now is how to model the behavior of the suture during manipulation by the robot and take that into account in the process of learning from demonstrations as well as incorporating the visual feedback in the process of tying the knot. This should greatly improve the robustness and applicability of our approach.

Goldman: Yes, we continue to explore the biological features of sand swimming, the physics of intrusion into granular media, as well as ways to improve robot performance. For example, expect papers on how the back muscles in the lizard are used during sand swimming, papers on lift control during sand swimming and papers on the physics of lift in granular media.

Sreetharan: This paper demonstrated the passive regulation of body-roll torques of an airborne robotic insect, largely resulting from aerodynamic drag. We expect to continue this research, demonstrating intelligent passive mechanisms that regulate a greater subset of forces and torques during flight. For example, a current project seeks to passively regulate yaw torques resulting from aerodynamic lift.

Furthermore, we plan to demonstrate active control under the PARITy methodology. Control inputs of this type do not alter wing trajectories, as per the conventional approach; rather, they bias the passive systems that regulate body forces and torques.

Did the group encounter any difficulties with team work? If so, how were these solved?

van den Berg: No, not really. The roles were clearly divided, and everyone was highly committed to the success of the project. Without this, it could not have succeeded.

Goldman: No, our team works great. My laboratory (we call it the CRAB Laboratory for Complex Rheology and Biomechanics) has a number of projects like this, in which physicists, biologists, and bioengineers work to gether to solve problems—such solutions in fact require the collaboration and skills from these different disciplines.

Sreetharan: Since this research was largely the result of individual effort, we had no major difficulties.

Thank you for the interview!

Academic Career Advice from Your Future Self

By Alejandro Perez

In robotics, real-time knowledge acquisition with no a priori data or "learning as you go" is very common, and it can be considered the norm. Similarly, as we develop our careers and grow as members of the academic community, we often find ourselves saying "If only I had known *x* two years ago?" Our field is rapidly growing, its rate of advancement is hastening, and it is slowly moving toward the spotlight of the entire scientific community. In the same way, joining top academic institutions or getting involved with cutting-edge research projects is getting more competitive each year. Below, you will find a short list of the most common "If only I had known's" I have heard from graduate students. Hopefully, they will help you through your career. Just consider it as advice from your future self.

Academic Research

A good transcript, graduate record examination score, and a statement of purpose are simply not enough anymore. Research is what will truly make you stand out and also what the bigger part of your graduate career will consist of. Many regret not getting involved with research from the very beginning.

Diversity

Most students get their first research experience at their own institution. However, getting results and good progress can tempt you to solely work with a certain laboratory. Working at just one place means meeting only a limited number of faculty members and having only one reference source and possibly a stale resume/CV. Consider working at your institution during the semester and applying for research internships or jobs at a different institution every summer. Some

Digital Object Identifier 10.1109/MRA.2010.940156
Date of publication: 14 April 2011

Robotics & Automation MAGAZINE | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

believe that it would have made their graduate school applications more attractive.

Publishing

Many consider that the best way to stand out is to at least have one publication early in your career. Publishing as an undergraduate or a young graduate is becoming fairly common. I've heard a number of students say that they could have published during the commencement of their careers if they had tried. Aim for it and do it. Use the hindsight to your advantage.

Graduate Courses

In most cases, a transcript with challenging graduate courses is more attractive than a generic transcript with perfect grades. It is difficult to measure one's ability to excel academically at a graduate school. Enrolling in these courses as an undergraduate will not only manifest this ability but also show your passion about the field. "I should have enrolled in that risky/interesting course?" is a common statement said while applying at a graduate school.

Every career is different; there is no clear black-and-white formula of what must be done to succeed. Consider these steps as an a priori probabilistic map of possible paths. As in robotics, it should be used as relevant data but not trusted as absolute and unchanging. Nevertheless, chances are that you will either develop a similar hindsight or look back in a few years and feel content for having listened to the advice from your future self.

The Reviewing Process: An Introduction for New Reviewers

Ludo Visser and Tamas Haidegger

To sustain a high-quality journal or conference, it is required that the submitted works are reviewed, meaning that the editor can properly decide whether a paper is suitable for publication. However, full-time reviewers are rare; therefore, a peer review was introduced: professionals review the papers of their peers (as volunteers), i.e., researchers that are active in the same field. To balance the system, it is required that each author participates actively. While paper reviewing is strongly connected to the academic career, people at industrial/government positions also need to review proposals, where similar rules apply.

It is clear that the success of the process depends on the quality of the reviews. Moreover, an additional outcome is that the authors receive a sincere and valuable opinion on their work that should help them to improve. As a reviewer, it is always important to deliver useful reviews, not in the least, because you will benefit yourself when you are an author or editor at a different occasion. But how to deliver a good review?

This article is the first in a new series supporting the RAS Student Reviewer Program (SRP, http://wiki.ieee-ras.org/ras_srp). The SRP aims at introducing young researchers to the reviewing process in a controlled and supervised way. A student's advisor should always be the first to assist, but with SRP, we hope to provide guidelines and a supporting infrastructure, in which the students receive structural feedback to improve themselves. This should help to improve both their reviewing and writing skills, and eventually the editorial boards and the professional community benefit from the high-quality reviews.

In these supporting articles, different aspects of the reviewing process will be highlighted, with the aim of providing guidelines for the young researchers new to the reviewing process. In particular, the series will address the following.

- How does the review process work and how do you get involved?
- How do you assess the quality and innovativeness of a paper?
- How do you evaluate the extent of readability and comprehensibility?
- How do you write a review that will help the authors to improve their work?
- How do you communicate your evaluation to the editors?

In this first article, we will provide a generic introduction to the reviewing process.

Digital Object Identifier 10.1109/MRA.2011.940899
Date of publication: 14 April 2011

The Reviewing Process

In a recent issue of *IEEE Robotics & Automation Magazine* (vol. 17, no. 4. pp. 101–104), Seth Hutchinson, editor-in-chief of *IEEE Transactions on Robotics*, already outlined the important parts of the reviewing process, with the aim of providing the authors some valuable insights. Here, we will do it from the reviewer's point of view.

Assignment of Reviewers

By the time you receive a request to review a paper, quite a few people have already looked at it. In general, the Editorial Board for a journal or a conference has two or three layers through which a paper advances before it reaches the reviewer. In the case of most IEEE publications, the paper is formally submitted to the editor-in-chief, who will assign it to an editor in

CALL FOR PAPERS

IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING

Special Issue on Automation in Green Manufacturing

There has been growing interest in green manufacturing worldwide, which has attracted substantial efforts from researchers in both academic and engineering communities. Green manufacturing deserves *efficient production of green technology products, and energy efficient and environmentally friendly manufacturing process and system*. To achieve this, automation is an essential component. In recent years, significant advancements in technology, the fast growing economy and rapidly changing market have generated numerous opportunities for innovation. At the same time, many new challenges have emerged in order to apply and implement these innovations. Such opportunities and challenges have substantially expanded the scope of automation. The goal of automation is to provide efficient scientific and engineering solutions for green manufacturing. The central theme of this Special Issue is ***emerging opportunities and future directions in automation for green manufacturing,*** where information technology based modeling, analysis, control and optimization are the focus areas. The purpose is to show the state-of-the-art research and applications in the general area of automation in green manufacturing, by bringing together researchers and practitioners from both academia and industry, to address the significant advancement, expose the unsolved challenges, present the needs for integration with new technologies, and provide visions for future research and development. This special issue aims to publish original, significant and visionary automation papers describing scientific methods and technologies with both solid theoretical development and practical importance. Submissions of scientific results from experts in academia and industry worldwide are strongly encouraged. Topics to be covered include, but are not limited to the following topics in green manufacturing,

- Design of equipment/robotics/processes for green tech products (solar panels, batteries, motors, wind turbines, fuel cell, etc.) manufacturing
- Automatic control in green tech products manufacturing
- Logistics, service and supply chain management of green tech products
- Factory modeling, analysis & evaluation in green tech products manufacturing
- Algorithms for planning, scheduling and coordination for green tech products manufacturing
- Process control and yield enhancement for green tech products manufacturing
- Mobile and wireless applications in green tech products manufacturing
- Inventory management for green tech products manufacturing
- Quality monitoring and control for green tech products manufacturing
- Efficient energy saving and management systems in manufacturing
- Emission/VOC reduction in manufacturing
- Real-time control to reduce energy consumption in manufacturing
- Automation and robotics for re-manufacturing
- Inventory control and scheduling for re-manufacturing
- Automation for waste management and recycling in manufacturing
- Lifecycle evaluation automation in manufacturing
- Planning and scheduling to optimize energy usage in manufacturing facilities
- Predictive maintenance and service for green manufacturing
- Manufacturing system design and operation for sustainability
- Integrated modeling of energy and environmental factors with production performance

GUEST EDITORS

Prof. Jingshan Li
Univ. of Wisconsin - Madison, USA
jingshan@engr.wisc.edu

Prof. James Morrison
KAIST, Korea
james.morrison@kaist.edu

Dr. Mike T. Zhang
Tianwei New Energy, China
mike.tao.zhang@gmail.com

Prof. Masaru Nakano
Keio Univ., Japan
m.nakano@sdm.keio.ac.jp

Dr. Stephan Biller
General Motors R&D, USA
stephan.biller@gm.com

Prof. Bengt Lennartson
Chalmers Univ., Sweden
bengt.lennartson@chalmers.se

IMPORTANT DATES

Aug 1, 2011: Submissions Due

Dec 31 2011: 1st Reviews Due

April 1 2012: 2nd Reviews

June 1 2012: Final MSS Due

Oct 2012: Tent. Publication

Paper Submission

All papers are to be submitted through the IEEE's **Manuscript Central** for Transactions on Automation Science and Engineering http://mc.manuscriptcentral.com/t-ase. Please select "Special Issue" under Manuscript Category of your submission. All manuscripts must be prepared according to the IEEE Transactions on Automation Science and Engineering publication guidelines http://www.ieee.org/publications_standards/publications/authors_journals.html. Please address inquiries to jingshan@engr.wisc.edu.

Digital Object Identifier 10.1109/MRA.2011.941110

charge of one of the subfields or topics that the journal or conference covers. The editor may then assign the paper to one of the (many) associate editors who are responsible for organizing the reviews. The (associate) editor contacts the reviewers (you) with an official request to review the paper. There are variations on this structure, with more or fewer layers, but the general idea is a pyramid structure.

Assuming the reviewer accepts the request and prepares the review on time, he/she submits it to the (associate) editor, who collects and summarizes all the reviews for that paper. The editor or editor-in-chief then makes a decision, which is communicated to the authors. This makes it immediately clear why you should deliver a good review: the editors need to integrate and summarize multiple reviews per paper, so it is important that they can easily understand your evaluation of the paper and quickly process it. Especially, in the case of large conferences, where the editors are faced with huge numbers of submissions and short deadlines, clear communication and efficiency is of great importance.

Let us discuss the process step by step. Potential reviewers are contacted because of their activity in a certain field. For the selection, the editor can rely on personal contacts but often refers to a list of people with assigned keywords, summarizing their expertise. Such lists are generally created automatically when online submission/review systems are used, such as Paper Plaza or Manuscript Central (these will be addressed in a future article). An individual enters the system when he/she submits a paper to a conference or journal, and the keywords chosen for the paper are also linked to the person. For example, assuming you submitted a paper with the keyword "humanoids" to a journal, the editor may now assume that you are active in the field of humanoids, and hence, you may be contacted with the request to review a paper that also has the keyword humanoids. Alternatively, the professors frequently delegate graduate students to provide an initial opinion on the paper they received to review.

Of course, being active in a field does not necessarily mean that you have sufficient expertise in the particular area that the paper addresses (e.g., if you have just started your program, your experience might be limited.) However, if you are contacted by an associate editor, it means that he or she has the opinion that your level of expertise is sufficient to review the paper, and thus you should, in principle, consider reviewing the paper. On the other hand, keywords do not tell the whole story of the paper, so it is possible that you get a request to review a paper in a field you know nothing of. To go back to the example, you may have written a paper on the mechatronics of walking humanoids, and thus chosen the keyword humanoids, while you may get a paper on the vision of humanoids. Then, of course, you should indicate you are not the right person to review that paper.

In summary, you will have to decide, for yourself, whether you accept a request to review a paper.

- In general, you should accept the request if you feel confident and can comply with the time schedule.
- Consider your choice based not only on the present knowledge you have but also on what you would like to learn. Reviewing a paper is a good way to broaden your scope.
- Be honest with the editor if you feel uncertain about your level of expertise.
- Reject a request if you know that you will not have enough time to deliver a thorough review.
- Also, reject a request if there is a conflict of interest, for example, if you have a professional or personal relationship with the author and arguably you cannot deliver an objective review.

If you decide to decline a review, the editor will highly appreciate it if you give the name of a person the editor might alternatively contact. Most importantly, whatever your decision is, reply promptly.

The Actual Reviewing

Assuming that you have accepted the request to review a paper, you are now officially a reviewer. It is now your responsibility to deliver a proper review on time. Requirements differ from conference to conference and from journal to journal. However, there are some common denominators.

First, it is important to take the process seriously. You should put a real effort into the review, as you would appreciate from other reviewers if you were the author. Also be considerate of the task of the associate editor(s). He or she will have to collect and summarize all the reviews and report to the editor. It is therefore important to be clear and concise in your review and to be on time. Especially, in big conferences such as ICRA, there is a huge number of reviews (several thousands) to be processed, and it is in everybody's interest to get through this process as smoothly as possible.

The SRP aims at introducing young researchers to the reviewing process in a controlled and supervised way.

First, read the paper again and again. This should be the second or third time you read it, since you should have read it before accepting the request to review. Read the paper carefully, make notes, and really try to understand its concept. Identify the strong and weak points, check the references if you are not familiar with them, and, most importantly, verify the claims the authors make. You will have to iterate this process a few times to form for yourself a clear opinion on the paper. This list is far from exhaustive but should give you some basis to start your review:

- Start with the title and abstract and determine whether they describe the content of the paper adequately. They should be self-contained. Specific length and format limitations usually apply.
- The introduction to the paper should outline the motivation for the work and the problem setting and present an overview of the relevant work on the topic. Preferably,

Robotics & Automation Magazine | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

an outline summarizing the paper should be given.

- The body of the paper is obviously the most important part. It must be clear from the text which part of the work is new and which parts rely on the previous work done by others. Pay attention to the technical details of the work but keep in mind that you are not required to redo the work of the authors. The presentation of the analysis, methods, and results is of great importance: the text should be readable (also for nonexperts to some extent), figures should be clear (check for legend, scale, and units), and the data should be presented concisely.
- The concluding section should summarize the work and not present any new information, nor should it just repeat the abstract. Any recommendations for future work should be relevant to the paper.
- The reference section is often overlooked despite its importance. It should contain relevant and recent work on the topic. A reference list dominated by previous works of the authors should raise your suspicion.

The central part of the review is your actual comments on the paper, and this part is most often forwarded to the authors.

Writing and Submitting Your Review

When you have formed your opinion on the work, it is time to communicate with the editor. At this point, the procedure will depend on the journal or conference you are reviewing for. In general, the editorial boards of IEEE conferences will ask you to rate the paper along a number of metrics (e.g., quality, innovation, readability, etc.), and in addition, ask you to provide comments to the authors. Optionally, you can provide private comments to the editor, which the authors will not see. Use this option if you have major concerns that may directly affect the overall outcome of the process. Journals may ask for specific information, especially, if the paper is a revision or resubmission.

In either case, it is important that your review is clear and concise, so the associate editor can quickly identify the main points of your concerns. For that reason, take care of the formatting of your review. (If there is a template provided, use it.) Moreover, especially if you are reviewing a journal paper, it is likely that you will be involved in a next iteration of the submission, i.e., a revised version. It is an advantage for yourself if you can easily read back your review to see what the concerns were with the previous version and to verify that those points have been addressed in the new version.

The central part of the review is your actual comments on the paper, and this part is most often forwarded to the authors (hence, make sure you do not reveal your identity if it is an anonymous review). It should be clear and constructive so that your points can be systematically addressed by the editor and authors.

- Begin with stating the title of the paper and provide a brief summary of the work. Not only does this help you in pinpointing the essence of the paper, but also help the editor, who will see many reviews and likely cannot remember each individual review.
- Outline why the paper is relevant and highlight the good points. This will result in a more balanced review. (There are very few really bad papers that are sent out for reviews.)
- Pointwise summarize your major concerns. Refer to the text so that both the editor and authors know what parts of the paper you are commenting on. Give advice on how the authors can address your concerns.
- Provide general recommendations to the authors to improve the paper (no paper is perfect). If you know of any relevant works, provide references that the authors can check. (Try to avoid referring only to your own publications.)
- Summarize minor concerns (optionally also in a bulleted list) and indicate how the authors should address them.
- In addition, you can provide a list of typographical and grammatical errors, and it will be highly appreciated. Make sure that the typo corrections are not the only useful advice you give.

Depending on the conference or journal, you may be asked to give a recommendation for the paper, e.g., "accept," "accept provisionally," "revise and resubmit," "reject," etc. This recommendation can either go in the review and be visible to the authors, or the recommendation is privately done to the editor.

In addition, you may need to write a cover letter to the editor. In this letter, you can privately reveal the concerns to the editor; for example, if you suspect plagiarism, or if you feel that there is something wrong with the paper that you cannot succinctly describe in the review. Also, your recommendation may be conveyed in the cover letter.

One thing that you must never do is to be insulting in your review. Even if you think the paper should be rejected, never use words like "terrible," or "ignorant." Such behavior is unprofessional. Remember, while the authors don't know you who are, the editor does, and the editor is not a person you want to think of you as unprofessional.

Further Reading

In this article, we briefly addressed the main phases in the reviewing process. Future articles in this series will enter into more details of each of this phase, and give more tips and guidelines that can help you. In the meantime, you can find more information on the Web site of the RAS SRP at http://wiki.ieee-ras.org/ras_srp, where you can contribute to the program and join our mailing list.

Acknowledgment

We are grateful to Dr. Manuel Ferre for his insightful comments on the SRP and for reviewing this article.

Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

Robotic Tools in Hospitals

By Raj Madhavan

Recognizing the labor-intensive and high employee turnover rate of sterile processing of surgical instruments in hospitals, Robotic Systems & Technologies, Inc. (RST), of Bronx, New York, thought that hospitals can benefit from lessons learned in manufacturing by using the tried and true automation techniques manufacturers employ to improve efficiency and quality. They saw a correlation between the repetitive nature of the sterile process department and assembly in the world of manufacturing. "When you think about it, the sterile processing department's tasks for the most part consist of counting, sorting, inspecting, and processing instruments just like an assembly line," said Dr. Michael R. Treat, an associate professor of clinical surgery in the College of Physicians and Surgeons of Columbia University, an attending surgeon at the New York-Presbyterian Hospital and founder of Robotic Systems & Technology, Inc. "Automation can perform these tasks precisely, automatically, and reliably 24 hours a day and seven days a week, allowing the administrators to cut costs, redeploy valuable employees to more challenging tasks, reduce errors, and ultimately protect patients."

The company developed the Penelope central supply (PenelopeCS) system as one of its first robotic tools because it hopes to be a line of multiple robotic applications addressing hospital supply chain needs. The PenelopeCS system will help automate key functions in the hospital's sterile supply department. The system uses RST's bagel software and vision system, and an Adept Viper six-axis robot from Adept Technology, Inc. The Adept Viper six-axis robot is a high-performance articulated robot designed specifically for precision applications. The robot is fittedwith a magnetic end effector that allows to pick up the instruments.

The hospital's sterile processing department has a dirty side where the instruments are washed down and disinfected and a clean side where they are prepped, packed, and sterilized in an autoclave. Working on the clean side of sterile processing, the PenelopeCS robot performs counting, sorting, and inspecting instruments. The robot ensures that each instrument tray sent to the operating room (OR) contains the correct instruments per the count sheet and that they are all in good working order. Furthermore, as these instruments are loaded, PenelopeCS will seamlessly update the hospital's inventory control system to provide traceability while reducing the workload on sterile supply staff.

The company's engineers have developed a robust and sophisticated robotic control language called *bagel*. The bagel software allows the PenelopeCS to formulate high-level goals like unload this tray of instruments or sort these instruments into stacks and to create a series of individual steps to accomplish those goals. If the operator interrupts PenelopeCS in the middle of this process, the system will remember the preempted goal and the last completed step so that it can resume the process after taking care of the operator's request. RST has also used bagel to capture an extensive knowledge base of information about surgical instruments: their types, shapes, synonyms, physical characteristics, and so on. PenelopeCS can use this knowledge base in conjunction with a proprietary machine vision system to identify surgical instruments. Third-party identification systems can also be integrated into the system to use barcodes to identify individual instruments.

"From our research, we estimate that hospitals adopting the PenelopeCS system will be able to improve quality, utilize labor more efficiently, and reduce the costs of processing surgical instrument trays by 23%," said Dr. Michael R. Treat. "We further predict that hospitals can expect a return on investment in less than 18 months." RST is working on potential hospital robotic applications in development, including an inspection application, to make sure that instruments open and close properly, making sure they are sharp and that they can cut. Other RST applications include an OR surgical assistant that can assist the surgeon by handing him/her the instruments he/she requires by utilizing a voice-recognition system. Currently, there are at least 5,800 hospitals in the United States alone that could benefit from this technology. A pilot program is currently in progress at a prominent New York hospital.

For further information, contact Robotic Systems & Technologies, Inc. (RST) at http://www.roboticsystech.com/ or e-mail mtreat@roboticsystech.com and Adept Technology, Inc. at http://www.adept.com or e-mail rush.laselle@adept.com.

Digital Object Identifier 10.1109/MRA.2011.940896
Date of publication: 14 April 2011

1070-9932/11/$26.00©2011 IEEE

Honors, Elections, and Other Member Activities

15 RAS Members Elected to the IEEE 2011 Fellows Class

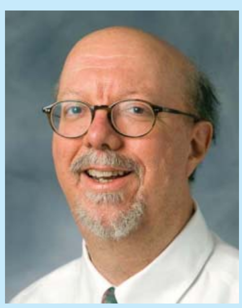
The IEEE Board of Directors has announced the names of the IEEE Members who have been elected as Fellows of the IEEE. The Members must be nominated to the grade of Fellow, and the nominations are carefully reviewed. No more than one tenth of 1% of the IEEE voting membership may be elevated to Fellow in a year.

New Fellows whose nominations were evaluated by the IEEE Robotics and Automation Society (RAS) include the following:

- Karl Bohringer, University of Washington—for contributions to microelectromechanical systems, parallel and distributed robotic manipulation, and self-assembly.
- Bruce Donald, Duke University—for contributions in robotics, microelectromechanical systems, and computational molecular biology.
- Pierre Dupont, Boston University—for contributions to modeling and control of frictional contact in robotics.
- Maja Matarić, University of Southern California—for contributions to robot coordination and learning in human–robot systems.
- Yoshihiko Nakamura, University of Tokyo, Japan—for contributions to robotics.
- Bradley Nelson, ETH Zurich, Switzerland—for contributions to nano- and microscale robots and systems.
- Allison Okamura, Johns Hopkins University—for contributions to the design and control of haptic systems and medical robotics.
- Leyuan Shi, University of Wisconsin-Madison—for contributions

2011 New Fellows of RAS

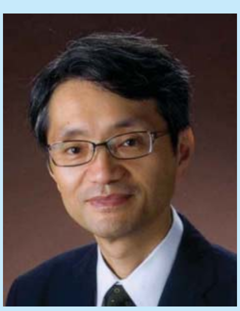
Karl Bohringer



Bruce Donald

Pierre Dupont

Maja Matarić



Yoshihiko Nakamura

Bradley Nelson

Allison Okamura

Leyuan Shi

Digital Object Identifier 10.1109/MRA.2010.940157
Date of publication: 14 April 2011

IEEE Robotics & Automation Magazine
Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

2011 New Fellows of RAS


Gaurav Sukhatme


Masaru Uchiyama


Manuela Veloso


Louis Whitcomb


Laurence Simar


Subhas Mukhopadhyay


Wen Jung Li

to nested partitions optimization methodology.

- Gaurav Sukhatme, University of Southern California—for contributions to multirobot systems.
- Masaru Uchiyama, Tohoku University, Japan—for contributions to design, modeling, and control of robotic structures.
- Manuela Veloso, Carnegie Mellon University—for contributions to the development of cognition, perception, and action in autonomous robot teams.
- Louis Whitcomb, Johns Hopkins University—for contributions to the theory and application of robotics for intervention in extreme environments.

The following RAS members were evaluated by other IEEE Societies and nominated to the grade of Fellow:

- Laurence Simar (IEEE Signal Processing Society), Rice University—for leadership in digital signal processor architecture development.
- Subhas Mukhopadhyay (IEEE Instrumentation and Measurement Society), Massey University, New Zealand—for the development of low-cost smart sensors and sensing systems.
- Wen Jung Li (IEEE Nanotechnology Society), Chinese University of Hong Kong SAR—for contributions in low-power integrated nanotube sensors and devices.

Other Honors

Tzyh Jong Tarn has been honored by the Chinese Academy of Sciences and the American Automatic Control Council.

The Chinese Academy of Sciences established the Einstein Professorship in 2005 and elects ten to 12 leading foreign scientists each year. . . . Recipients of the award till now include six Nobel Prize laureates in fields, ranging from economics to physics, and one Turing Award winner.

This year, they elected 11 recipients including Tarn, who is the first engineer to receive the award. The American Automatic Control Council has awarded Tarn the 2010 John R. Ragazzini Award for substantial contributions to control education through teaching, mentoring of graduate students, and research in control theory and applications to robotics.

Ren-Chyuan Luo of National Chung Cheng University, Taiwan, has been awarded the Harashima Award for innovative technologies for his technical contributions in intelligent mobile robotics and his leadership in International Conference on Intelligent Robots and Systems (IROS).

RAS Members Elect New AdCom Members

Congratulations and thanks to the following RAS members who were elected to serve on the Administrative

RAS New AdCom Members

Martin Buss

Toshio Fukuda

Bradley J. Nelson

Lynne E. Parker

Cecilia Laschi

Peter Ian Corke

Shigeki Sugano

Committee (AdCom) for a three-year term beginning in January 2011:

- Toshio Fukuda, Nagoya University, Japan
- Bradley J. Nelson, ETH, Zurich, Switzerland
- Lynne E. Parker, University of Tennesee-Knoxville
- Cecilia Laschi, Scuola Superiore Sant'Anna, Italy
- Peter Ian Corke, Queensland University of Technology, Australia
- Shigeki Sugano, Waseda University, Japan.

In addition, President Kazuhiro Kosuge appointed Martin Buss of the Technical University of Munich to fill the unexpired term vacated by Roland

Call for IEEE RAS AdCom Nominations for 2011 Election

The RAS membership will elect six new members of the AdCom in 2011, each to serve a three-year term beginning in January 2012. The AdCom is the governing body of our Society.

Responsibilities of AdCom Members

The AdCom members must attend two formal meetings each year, one in conjunction with the *IEEE International Conference on Robotics and Automation* and the other usually in October/November and, usually, in conjunction with another major conference. Each Adcom member is expected to serve on at least two of the major boards and/or committees of the Society.

Eligibility

Any higher-grade member of the Society is eligible to serve, and all higher-grade members as well as graduate students may nominate the candidates and vote. To nominate a candidate or offer yourself as a candidate, contact the Society administrator (r.g.snyder@ieee.org) by 15 June 2011.

Candidates may also petition to be on the ballot. All persons who, by the deadline, submit petitions with valid signatures and IEEE Member numbers with at least 2% of the year-end voting membership will be placed on the ballot (160 signatures of eligible voters).

Persons submitting petitions with at least 25 valid signatures and IEEE Member numbers will be considered by the Nominating Committee. Only original signatures on the paper or electronic signatures submitted through the RAS Web site will be accepted. Faxed or e-mailed signatures are not acceptable. Contact the Society administrator, Rosalyn Snyder (r.g.snyder@ieee.org), to obtain a paper petition form or set up an electronic petition. Paper petitions with signatures must be submitted by 15 May 2011 to be placed on the ballot.

The Nominations Committee will consider all nominations and petitions and select the candidates to be placed on the ballot.

Special issue on
Robotics for Environmental Monitoring

A special issue of the ***IEEE Robotics and Automation Magazine***

Robotic systems are increasingly being utilized as fundamental data gathering tools by scientists allowing new perspectives and greater understanding of the planet and its environmental processes. Today's robots are already exploring our deep oceans, tracking harmful algal blooms and pollution spread, monitoring climate variables, and even studying remote volcanoes.

With increased awareness of the ability of robots to collect scientifically relavant information, new opportunities are arising for large-scale environmental monitoring that will push the frontiers of robotic and natural sciences. Addressing these opportunities will present significant challenges in field robotics research, stimulating new findings in the direction of robotic systems more able to perceive, plan, move and actuate in natural environments.

These new environmental monitoring robots with increased autonomy, endurance, planning and sensing capabilities will be expected to interact in teams and within other data gathering networks for efficient and precise measurement of environmental processes at scales never before seen. This special issue in Robotics for Environmental Monitoring seeks contributions on recent developments and applications in key areas of the advancing field of environmental robotics.

Scope, description and more information

Topics of interest include but are not limited to:

- Robotic systems for observation of natural environments, being them ground surfaces, air, water surface, underwater, or underground
- Robotic systems able to track natural phenomena
- Sensors for environmental monitoring
- Mobile sensor networks
- Spatio-temporal data processing and environment modeling
- Low energy design
- Energy harvesting
- Other subjects relevant to robotics for environmental monitoring

GUEST EDITORS		Important Dates	
Lino Marques Universidade De Coimbra, Inst. Sistemas E Robotica *Coimbra, Portugal* lino@deec.uc.pt	Matthew Dunabin CSIRO Australia Matthew.Dunbabin@csiro.org	1 July 2011 1 November 2011 1 December 2011 March 2012	Submissions Due First Review Second Review Publications

Digital Object Identifier 10.1109/MRA.2011.941109

Siegwart, who was unable to complete his term due to his professional commitments.

Nominations for the 2011 AdCom elections should be sent to the RAS administrator by 15 June 2011 (see "Call for IEEE RAS AdCom Nominations for 2011 Election").

RAS Membership Activities Board News

Vice President Stefano Stramigioli reports that the Member Activities Board (MAB) has started working on a new metamodel and is in the process of defining the new activities for the various committees. The board needs to concentrate on the following:

1) creating new services and advantages for the membership
2) improving the awareness of such services and advantages
3) capitalize on the good work and extend membership and involvement in the Society.

On the member services side, we are considering a better definition for a mission of the Gold Lunch within the new context of MAB, and we have now a structural budget for a lunch organized in cooperation with the Women in Engineering. Among the new ideas, which are in the air, is an extended submission deadline for conferences for members, early access of portable document format programs for conferences, an e-book program,

IEEE Robotics & Automation Magazine 2010 Reviewers

We recognize the invaluable contributions of the following individuals who served as *IEEE Robotics & Automation Magazine* reviewers in 2010. Without their dedication and hard work, this magazine and other technical publications could not exist. Thanks to our reviewers!

Artur Abdullin
Nawaf Ali
Dimitris Alimisis
Konstantinos Andrianesis
Gianluca Antonelli
Filippo Arrichiello
Jonathan Becedas
Pierluigi Beomonte Zobel
Fabio Paolo Bonsignorio
Pieter Cornelis Breedveld
Adrien Briod
Terry Bynum
Darwin G. Caldwell
Sylvain Calinon
Domenico Campolo
Rafael Capurro
Raffaella Carloni
Alicia Casals
Julian Castellanos-Ramos
Daniela Cerqui
Rawichote Chalodhorn
Lyle Chamberlain
Raja Chatila
Antonio Chella
Thomas Christaller
Henrik Iskov Christensen
Mark Coeckelbergh
A. Paulo Coimbra
Roberto Cordeschi
Peter Corke
Torbjorn Dahl
Karthik Dantu
Prithviraj Dasgupta
Edoardo Datteri
Francisco De Leon Gomez Maqueo
Michael Decker
Emel Demircan
Jeanne Dietsch
Nicola Diolaiti
Evan Drumwright
Amy Eguchi
Charles Ess
Juan-Antonio Fernandez-Madrigal
João Filipe Ferreira
Douglas Few
Paolo Fiorini
Maria Teresa Francomano
Dimitri Gavrilov
Marina Gavrilova
Eugenio Guglielmelli
Stephan Guttowski
Raia Hadsell
Norihiro Hagita
Tatsuya Harada
Alwin Hoffmann
Geoffrey Hollinger
Stefan Hrabar
Loulin Huang
Odest Chadwicke Jenkins
Nikola Jetchev
Jonathan Kelly
Farid Kendoul
Volker Krueger
Norbert Krüger
Haruhisa Kurokawa
Piero Larizza
Cecilia Laschi
Nicola Lettieri
Weiting Liu
Giuseppe Longo
Damian Lyons
Bruce Macdonald
Karon Maclean
Paola Maestro
Patrick Martin
Luis Mejias
Silvestro Micera
Maruf Monwar
James Moor
Giovanna Morgavi
Jun Morimoto
Shinńchiro Nakaoka
Lorenzo Natale
José Neira
Tatsuya Nomura
Edwin Olson
Ming Ouyang
Erhan Oztop
Lynne Parker
Davide Parmigiani
Sujit Pb
Arvind Pereira
Kenneth Pimple
Carlo Pinciroli
Christian Plagemann
Frederick Norman Pollard
Andreas Pott
Cedric Pradalier
Terenziano Raparelli
Leon Reznik
John Rieffel
Laurel D. Riek
Jennifer Robertson
Ferdinando Rodriguez y Baena
Septimiu E. Salcudean
Jelle Saldien
Tobias Seidl
Andrei Sherstyuk
Takanori Shibata
Bruno Siciliano
Richard Simpson
Sanjiv Singh
Elizabeth Sklar
Ryan Smith
Paolo Soda
John Spletzer
Dimitar Stefanov
Timothy Stirling
Salah Sukkarieh
John Sullins
Adnan Tahirovic
Ken Taylor
Franco Tecchia
Russ Tedrake
Eduardo Torres-Jara
Marc Toussaint
Ali Emre Turgut
Ales Ude
Barkan Ugurlu
Ravi Vaidyanathan
Peter Van Lith
Bram Vanderborght
Sethu Vijayakumar
Wendell Wallach
Hanlei Wang
Jens Wawerla
Jutta Weber
Thomas Wimboeck
Martijn Wisse
Shandong Wu
Yoji Yamada
Roman Yampolskiy
Eugenio Yime
Tomoaki Yoshikai
Ge Yunjian
Yuru Zhang
Chunlin Zhou

Digital Object Identifier 10.1109/MRA.2010.940159

Robotics & Automation MAGAZINE | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

and the establishment of a Robotics Hall of Fame. Membership is doing very well, and we have a significant positive trend since 2006. A new procedure is under investigation for improving the awareness among the members on the sponsoring that we give to Chapters and making the international activities more remarkable. This will require a clear commitment from our Chapters as well. On the education side, we are structuring competitions by trying to create a standing committee on competitions and to work toward a recognized summer school program.

Our student committee is extremely active; for the first time, a separate Student Activities Committee (SAC) Meeting was held at IROS 2010 to facilitate the organization of future SAC activities. As a major step forward in recruiting, SAC prepared a brochure and flyer to promote RAS SAC. (These are available at SAC page: http://wiki.ieee-ras.org/mab/sac.)

Chapter News

Madras India

A total of 85 engineers, research scientists, and consultants in India, IEEE Madras section, participated in the program on 30 October 2010 to inaugurate the new RAS Section Chapter. The meeting was arranged in association with M/s. Larsen & Toubro Ltd., ECC Division, Chennai. The RAS Chapter was formally inaugurated by Dr. T. Thyagarajan, chair, IEEE Madras Section. Dr. G.V. Rao, chair, RAS (Madras Chapter), gave an excellent insight into robots and their applications in various areas and also briefly highlighted the history of the evolution of robots. Other participants in the program were Er. S. Rajavel, chair, IEEE Industry Applications Society (Madras Chapter), executive vice president, M/s. Larsen & Toubro Ltd., Chennai, and the Program Speaker Er. S. Malakar, head IT Systems, M/s. Larsen & Toubro Ltd., Delhi International Airport Ltd.

Southern Alberta Chapter

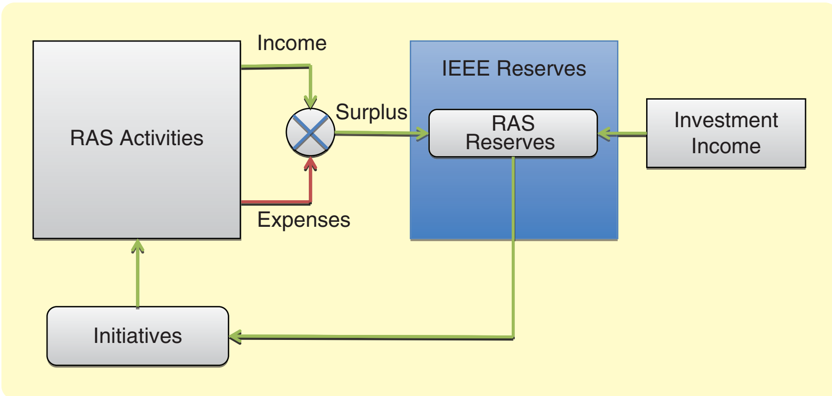
The IEEE/Western Canadian Robot Society (WCRS) Games took place on 15 May 2010 at the Calgary Aerospace Museum in their main display hanger. The event was coorganized by Southern Alberta IEEE and WCRS. Detailed information can be found at www.robotgames.com/2010-robot-games/. The RAS MAB provided a Chapter Development Grant for US$2,000, which covered the audiovisual for the event, consisting of a live camera fed to giant video screens and sound system for the MC/play-by-play person. Other funds and in-kind services were supplied by the local IEEE Southern Alberta Section and other sponsors.

More than 80 competitors and 200 spectators came for the games, which started at 10:00 a.m. and lasted till 4:00 p.m. Television media from a local station covered the event.

The traditional competition events carried over from previous years were Art-Bot, Mine-Sweeper, Line-Follower, BEAM Walker, Mini-Sumo, and Full-Size Sumo and included 41 competitors from all walks of life: high-school students, university students, and amateur roboticists.

A new competition was included this year: Lego Mindstorms Treasure Hunt. A University of Calgary graduate student of RAS Chapter chair Chris Macnab was sponsored by an Imperial Oil Inspiring Careers in Science/Technology/Engineering/Mathematics (STEM) Grant to mentor elementary school girls to build Lego robots. She organized this competition event for them so that they could compete their robots by the end of the year, and 44 students came to compete their robots.

Nominations for the 2011 AdCom elections should be sent to the RAS administrator by 15 June 2011.



Although the event was not restricted to girls, the prize-winning teams ended up being exclusively all-girl teams. The girls and their teachers assured the organizers that we had, in fact, interested the girls in considering careers in STEM. The girls completed the surveys for their attitudes toward careers in STEM at the beginning of the program and after the games, and this data will be used in a paper for the American Society for Engineering Education Annual Conference in 2011.

Two guest speakers gave their presentation in a conference room to create an intimate setting, but the audio–video was broadcasted to the entire (hanger) space during the lunch break. Carlo Menon from Simon Fraser University (www.ensc.sfu.ca/~cmenon/) spoke about his group's effort to build a wall-climbing hexapod robot using a gecko-inspired design. Then, Mahdi Tavakoli from the University of Alberta (people.seas.harvard.edu/~tavakoli/) spoke about the latest developments in surgical robotics and haptics research. Both speeches were received with enthusiasm by the audience.

Deadline for RAS Local Chapter Development Grants

The RAS MAB awards a limited number of Chapter Development Grants to local RAS Chapters for professional development, educational outreach, and other programs. Grant proposals will be reviewed by MAB at their meeting in Shanghai in May 2011 and funds up to US$2,000 will be awarded on a competitive basis. The deadline for proposals is 30 April 2011. For submission details, see http://www.ieee-ras.org/chapters.

Automation Forum at ICRA

May 10 (Tuesday), 2011
Shanghai International Convention Center
(SHICC)

- Organizer: **Dr. Peter B. Luh**, University of Connecticut, USA

This forum introduces the latest trends in automation, including life science, business and software, semiconductor, transportation, and control and optimization. Panelists are:

- **Dr. Tianwei Jing**, Director of R&D, Nano Measurements Division, Agilent Technologies, Inc, USA
- **Mr. S. V. Subrahmanya**, Senior Vice President, Infosys
- **Dr. John Du**, Director, China Science Lab, GM China, China
- **Dr. Joseph Zhifeng Xie**, Chief Executive Officer, Advanced Semiconductor Manufacturing Corporation (ASMC), China
- **Dr. Jian Chu**, Chairman of the Board, Supcon, Professor of Institute of Cyber-Systems and Control (CSC) and Vice-President, Zhejiang University, China
- http://www.ieee-ras.org/news/show/id/276

Digital Object Identifier 10.1109/MRA.2011.941108

ERRATA

In the article "The SMACH High-Level Executive" of the "ROS Topics" column of the December 2010 issue of *IEEE Robotics & Automation Magazine* [1], the name of one of the authors was misspelled. The correct name of the author is Jonathan Bohren. We apologize for the error.

Reference

[1] J. Boren and S. Cousins, "The SMACH high-level executive," *IEEE Robot. Automat. Mag.*, vol. 17, no. 4, pp. 18–20, Dec. 2010.

Digital Object Identifier 10.1109/MRA.2011.940895
Date of publication: 14 April 2011

2011

9–14 May
ICRA 2011: IEEE International Conference on Robotics and Automation. Shanghai, China. http://2011.ieee-icra.org/

20–23 June
ICAR 2011: 15th International Conference on Advanced Robotics. Tallinn, Estonia. ICAR. http://www.icar2011.org

20–24 June
Med 2011: 9th Mediterranean Conference on Control and Automation. Corfu, Greece. http://www.med2011.org/

21–24 June
WCICA 2011: 9th World Congress on Intelligent Control and Automation. http://www.wcica2011.org.tw/about.php

29 June–1 July
ICORR 2011: IEEE 12th International Conference on Rehabilitative Robotics. Zurich, Switzerland. http://www.icorr2011.org

28 July–1 August
ICINCO 2011: 8th International Conference on Informatics in Control, Automation and Robotics. Noordwijkerhout, The Netherlands. http://www.icinco.org/

4–6 July
AIM 2011: IEEE/ASME International Conference on Advanced Intelligent Mechatronics. Budapest, Hungary. http://www.aim2011.org/

7–10 Aug.
ICMA 2011: IEEE International Conference on Mechatronics and Automation. Beijing, China. http://2011.ieee-icma.org

22–26 Aug.
MMAR 2011: 16th International Conference on Methods and Models in Automation and Robotics. Miedzyzdroje, Poland. http://www.mmar.edu.pl/

24–27 Aug.
IEEE-CASE 2011: IEEE Conference on Automation Science and Engineering. Trieste, Italy. http://www.deei.units.it/CASE2011/.

25–30 Sept.
IROS 2011: IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco CA USA. http://www.iros2011.org.

2–4 Oct.
ARSO 2011: IEEE Workshop on Advanced Robotics and Its Social Impacts. San Francisco Bay Area USA. Submissions due 15 April. http://www.arso2011.org/

27-29 Oct
Humanoids 2011: 11th IEEE-RAS International Conference on Humanoid Robots. Bled, Slovenia. http://www.humanoids2011.org

Digital Object Identifier 10.1109/MRA.2010.940162
Date of publication: 14 April 2011

IEEE ROBOTICS & AUTOMATION SOCIETY

Digital Object Identifier 10.1109/MRA.2011.941112

Turning Point *(continued from page 112)*

Figure 2. Eric Paulos with personal roving presence (PRoP) in 1998. (Photo courtesy of Eric Paulos.)

One of the benefits of telepresence robots is that they enhance your sense of agency in the remote environment. You are not just a passive conversationalist; you can take the initiative to move around and explore. That really matters. It allows for spontaneity and greater potential for discovery.

EG: **In 2001, you edited a collection of essays titled, *The Robot in the Garden: Telerobotics and Telepistemology in the Age of the Internet* [5]. One of the articles, by John Canny and Eric Paulos (Figure 2), describes a telepresence robot [6] (Figure 3) very similar to the commercial versions we're seeing today. Why did it take nearly ten years for these robots to become commercially viable?**

One of the benefits of telepresence robots is that they enhance your sense of agency in the remote environment.

KG: When people like John Canny and Eric Paulos were developing telerobots and camera systems, the Internet and wireless networks weren't as fast and reliable as they are today. Now networks have more bandwidth and better quality of service. That makes a huge difference. The other thing that has changed is that it's less expensive to build a robot today, because the components you need are getting better and cheaper. So companies like Vgo, Anybots, and Willow Garage are now commercializing these robots. When can they get the price down to a point where it's available to a large number of people? When that happens, things will get very interesting.

EG: **Today we use cell phones, e-mail, instant messaging, Twitter, Facebook—and soon some of us may be using telepresence robots. We're staying connected in more ways and for longer periods of time. Where is this going?**

KG: Last spring I taught a course with UC Berkeley philosopher Hubert Dreyfus on the philosophy of technology. Our aim was to give students, many of whom will be creators of technology, a broader historical and social perspective to understand technology. Our starting point was the 1954 essay by Heidegger, "The Question Concerning Technology" [7]. Let me say first that Heidegger is a problematic figure. He was deeply flawed personally. But we can't dismiss everything he wrote. He's one of the most influential philosophers of the 20th century.

In a nutshell, Heidegger asserts that technology is not specific tools or methods. "The essence of technology is nothing technological." Instead, technology is a mode of being, a philosophical attitude that we're immersed in. This is not something we have consciously adopted; it's all around us, we're engulfed in it. Heidegger calls this *Gestell* and argues that the essence of this technological mode of being is a drive to make the world increasingly available for use.

For example he considers the Rhine River. Rather than approaching the river as primitives, who ponder how the gods created it, or as poets, who focus on its unique beauty, we approach the river as a resource available to generate power. We build coal and oil stockpiles, vast hedge funds, and comprehensive databases and cloud networks. We treat the world as a resource and want to make it more and more accessible for future use. The most popular technologies of our age are those characterized by flexibility and their ability to be reconfigured, such as polymers, genomics, stem cells, nanotechnology, the Internet . . . and robots.

Heidegger warned that this worldview can overwhelm us: we'll start applying this attitude to ourselves. We'll view ourselves as resources, and make ourselves increasingly available. This seems to be coming true with cellphones and laptops and Facebook

Figure 3. Anybot commercial telepresence robot (2010).

Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

and Twitter: compared with ten years ago, we're developing an overwhelming personal sense of obligation to be constantly online, exposed, and available.

EG: I guess Heidegger would have hated telepresence robots . . .

KG: Exactly. Telepresence makes the world more available. As researchers, we're excited about it, but from Heidegger's perspective, it's another step along a dangerous trajectory. At the end of the essay, he says we're at a crossroads, we can continue toward a supreme danger, where we are engulfed and overwhelmed and transform ourselves into resources. But Heidegger also saw a bright side, a way out of this situation. As this trend continues, maybe we'll be jolted into realizing what we're doing and develop the capability to resist it, to set boundaries. In other words, maybe we have to hit bottom before we can stop the madness.

EG: Almost ten years ago, you led a telepresence research project called the *TeleActor* [8], using people as proxies for other people (Figure 4). Is the Tele-Actor a precursor to robotic telepresence?

KG: Our idea was to hire an outgoing person—a Tele-Actor—who could go to a place you're unable to go yourself. The Tele-Actor would wear a camera and microphone, and you'd see and hear as though you were there. The idea was to allow large groups of students or citizens to share remote experiences. For example, allowing a group of disadvantaged students to collaboratively steer a Tele-Actor through a working steelmill in Japan, visit a working microelectronics facility, or attend a dinner at the White House . . .

Dez Song and I were awarded an NSF grant to study collaborative telerobotics. We could have used a robot, but we needed a highly agile, adaptable, and outgoing agent. We joked that rather than robots replacing people we'd have a person replace the robot. We did a lot of experiments, but the technology was not there yet. In 1999, we started with analog video, and we were constantly getting interference. Then we switched to an early version of WiFi, and network connections were slow and unreliable. On the client side, we prototyped using Java applets. It was primitive. I wished we had 4G networks back then.

Figure 4. Tele-Actor Annamarie Ho. (Photo courtesy of Bart Nagel.)

EG: So when the operator spoke, the Tele-Actor repeated what was said?

KG: The key idea was that there would be more than one operator. Think of an actor taking directions from a group of remote directors. The Tele-Actor has to improvise. We investigated the interactions that would take place and see how they'd compare to normal situations. I think someone should repeat the experiment today.

EG: And in the future we can replace Tele-Actors with androids! We're already seeing some steps in that direction. What do you think of telepresence robots that look like people, like the androids Hiroshi Ishiguro is creating [9]?

KG: Hiroshi's robots are not only very human but also very specific to individual humans. They are designed to act as surrogates in a very real way. This work has connections to psychology, mythology, and science fiction. It goes back to Galatea, the Golem, and later Pinocchio and Frankenstein and Blade Runner, and all the attempts to create something that's very lifelike. Hiroshi is pushing the limits and asking deep questions about how we view ourselves such as the Cartesian question: Are we automatons or not? [10] Androids and humanoids can help explore these questions. Maybe we'll discover things we want to avoid [11]. The only way we're going to find out is by experimenting.

References

[1] IEEE Spectrum's telepresence special report [Online]. Available: http://spectrum.ieee.org/static/telepresence

[2] CONE Project [Online]. Available: http://www.c-o-n-e.org/

[3] Steering flexible needles through soft tissue [Online]. Available: http://automation.berkeley.edu/projects/needlesteering/

[4] Available: http://www.tfcbooks.com/special/missiles.htm

[5] Available: http://goldberg.berkeley.edu/art/tele/

[6] Personal Roving Presence Project [Online]. Available: http://www.prop.org/

[7] M. Heidegger, *The Question Concerning Technology Other Essays*. Harper Torchbooks, Jan. 19, 1982

[8] Tele-Actor Project [Online]. Available: http://teleactor.berkeley.edu/

[9] IEEE Spectrum articles on Hiroshi Ishiguro [Online]. Available: http://spectrum.ieee.org/tag/hiroshi+ishiguro

[10] Descartes' meditations [Online]. Available: http://www.wright.edu/cola/descartes/

[11] K. Goldberg, Ed. *The Robot in the Garden: Telerobotics and Telepistemology in the Age of the Internet*, Cambridge, MA: MIT Press, June 2000.

Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

IEEE RA

Photo Credit: Bruce Kluckhohn

Honorary Chairs Committee
George Bekey
Kazuhiro Kosuge
C.S. George Lee
Bruno Siciliano
T.J. Tarn

General Conference Chair
Nikos Papanikolopoulos

General Conference Co-Chairs
Henrik Christensen
Shigeki Sugano

Program Chair
Paul Oh

Program Co-Chairs
Kostas Kyriakopoulos
Stergios Roumeliotis
Satoshi Tadokoro

Finance
Venkat Krovi

Videos
Stefano Stramigioli

Workshops/Distinguished Lectures
Gaurav Sukhatme

Tutorials
John Spletzer

Registration
Herbert Tanner

Publications
Max Meng

Exhibits
Raj Madhavan

Publicity
Maria Gini

Awards
Danica Kragic
Kevin Lynch
Frank Park

Robotics Challenges
Steve Cousins

Call For Papers

ICRA is the robotics community's premiere annual academic conference and will be held in the Saint Paul, MN on May 14-18, 2012. The ICRA theme is "**Robots and Automation: Innovation for Tomorrow's Needs.**" Robotics and automation are at the crossroads of new developments in algorithms, hardware, and software that pave new routes in technological innovation.

Contributed Papers

Prospective authors should submit PDF versions of their paper. Six pages in standard ICRA format are allowed for each paper, including figures. A maximum of two additional pages is permitted. Detailed instructions for submission are available on the conference website.

Invited Sessions

Invited sessions on new topics or innovative applications will be considered. These sessions, subject to the regular review process, will consist of four to six papers. Prospective organizers should include a brief statement of purpose for the sessions as well as the abstracts of the papers to be included.

Videos

1.5 to 3 minute videos illustrating new and exciting results will be showcased in a special video session. Submission instructions are detailed on the conference website.

Tutorials and Workshops

Proposals for half- or full-day tutorials and workshops will be reviewed according to instructions on the conference website. The Tutorials Chair will actively recruit tutorial organizers to create theme-based archival material. It is expected that the Technical Committees will play a more active role in the workshops. Prospective organizers should work with the appropriate Technical Committee to propose relevant topics.

Exhibits

Both physical and virtual exhibit sites will be featured at the venue. Information for virtual exhibitors will be available on the web site.

Important Dates

September 16, 2011: Submission deadline
January 6, 2012: Acceptance notification
February 10, 2012: Final contribution deadline

URL: http://www.icra2012.org

Digital Object Identifier 10.1109/MRA.2011.941107

AD INDEX

The Advertisers Index contained in this issue is compiled as a service to our readers and advertisers: the publisher is not liable for errors or omissions although every efforts is made to ensure its accuracy. Be sure to let our advertisers know you found them through *IEEE Robotics & Automation Magazine*.

Digital Object Identifier 10.1109/MRA.2011.940894

AD SALES OFFICES

Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

TURNING POINT

Telerobots

By Erico Guizzo

In this issue, Erico Guizzo (EG) interviews Ken Goldberg (KG), IEEE Fellow and engineering professor at the University of California, Berkeley, about telerobots, androids, and Heidegger.

EG: In the past year, six companies started selling telepresence robots [1]. I tested two of the robots myself, discussing at length their technical merits as well as their practical shortcomings. Telepresence robots didn't come out of nowhere; they stem from a convergence of different technologies, each with its own history. The advent of robotic telepresence also reflects a powerful trend where many of us are becoming ever more connected and available. What made these robots possible now? What's so appealing about roaming around as a machine in a remote place? And where is this technology taking us, literally and figuratively?

To explore these themes, I spoke with Ken Goldberg (Figure 1), IEEE Fellow and engineering professor at the University of California, Berkeley, and a member of *IEEE Spectrum*'s editorial advisory board. When he's not developing geometric algorithms for automation or prototyping robot cameras to spot wild birds [2] or computer-controlled flexible needles that steer through soft tissues [3], he's delving into the interactions between technology, art, and media. If anyone can make a connection between robots and Heidegger, it's Ken.

EG: I recently asked Marvin Minsky what he thought of current telepresence robots. He complained that they don't have legs. And I've seen other people complaining that they don't have arms. What do you think of their design?

Figure 1. Ken Goldberg served as a RAS Vice President of Technical Activities from 2006 to 2009. He studied algorithmic automation, medical robotics, and networked telerobotics. He also cotaught a course at UC Berkeley on the philosophy of technology. (Photo courtesy of Kathrin Miller.)

KG: Wheels are probably sufficient. When you add arms and hands you need more actuators, more sensors; it increases costs. But robotic parts and technologies are getting better and less expensive. Brian Carlisle [former CEO of Adept Technology] observed that we can buy a car, which includes a ton of metal and is filled with actuators and sensors, for under US$10,000: we should be able to do the same for robots. Volume reduces cost. People want robots that clean the house (and change diapers) but that will take much more research. In the meantime, can robots enhance communication?

That's the idea behind the new generation of telepresence robots. They build on infrastructure such as fourth generation (4G) and wireless fidelity (WiFi), but how should they be designed? For instance, the Rovio [a home robot sold by WowWee] is about the size of a cat, so you can't have an eye-to-eye conversation with a human, unless you want to talk to your cat. Eye contact is important. So is the ability to point to things in the environment, which can be accomplished with a laser and a two-axis gimbal. There are many design issues in making telepresence a compelling experience.

EG: And why do we want to physically extend ourselves to distant places anyway? Telephone and Skype aren't enough?

KG: The idea of remote control, you click a button here and something happens over there, is a very powerful and satisfying experience. We love our TV and garage remotes. The history of robots is intertwined with the history of remote control. It goes back to Tesla's experiments with a radio-controlled boat, which he demonstrated in New York in 1898 [4]. After World War II, the first robots were master-slave telerobots used to handle radioactive substances. Today, telerobots are used for exploration, in space and underwater, and for bomb disposal. Telepresence is different because you're not manipulating an object or performing a repair; you're interacting with people. There are humans on both ends. The goal is to give the remote operator a sense that he or she is closer to the people on the other end. And hopefully vice versa.

(continued on page 108)

Digital Object Identifier 10.1109/MRA.2010.940163
Date of publication: 14 April 2011

Robotics & Automation Magazine | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

WAM™
whole arm manipulator

TRUE HAPTICS | BACKDRIVABLE | SAFETY LAYERED | UNIVERSAL | MOBILE READY | FEATHER TOUCH

reach **1 m** weight **27 kg**

design :: Barrett Communications :: www.barrettcommunications.com

The **Standard** in Advanced Manipulation

Barrett Technology® Inc.

The WAM™ arm and BarrettHand™ manipulators represent the new standards for advanced robotics research with unmatched versatility. Visit Barrett's website at **http://www.barrett.com** to learn more.