FROM THE FIELD

Field-Testing Astronaut Assistance Robots in Australian Outback

By Graham Mann, Nicolas Small, Kevin Lee, Jonathan Clarke, and Raymond Sheh

The trouble with field-testing robots is that we are taking complex machines out of the laboratory and into the dirt: natural, unstructured environments that cannot be easily characterized or measured. There they could be doing imperfectly characterized tasks. We expect robots to be behaviorally flexible so describing a typical task will generally underspecify actual usage. The machine design, task, and environment are not orthogonal factors either, since they might interact in complicated ways. As if all this was not enough, most field robots are still teleoperated, which adds the attendant problems of evaluating the human controller and interface. Published work in this area tends to focus on demonstrating the robot's fitness for purpose based on specific requirements, often according to the contingencies of practical funding. Too often that commits the work to studies of performance on tasks that are not necessarily well understood, or even particularly well described, and to measurements within environments that cannot be duplicated.

The machine design, task, and environment are not orthogonal factors either, since they might interact in complicated ways.

Digital Object Identifier 10.1109/MRA.2015.2452200
Date of publication: 11 September 2015

Figure 1. A simulated Mars walk. (Photo courtesy of G. Mann.)

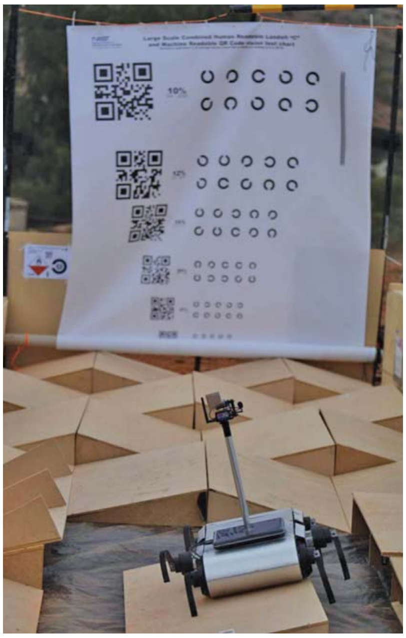

Figure 2. The hexapod vision test. (Photo courtesy of N. Small.)

How can we put field-testing on a more scientific footing? What we need is practical and widely accepted standards for robot testing, followed up by excellent sharing of results and an honest comparison of performance as a function of design. A step in the right direction is the U.S. Department of Homeland Security–National Institute of Standards and Technology–American Society for Testing and Materials (DHC–NIST–ASTM) tests for emergency-response robots. A lot of effort has gone into creating and documenting tens of useful test rigs and task score sheets, which the general community can easily build and use (www.nist.gov/el/isd/ks/upload/DHS_NIST_ASTM_ Robot_Test_Methods-2.pdf).

(continued on page 191)

2015

14–15 September
MFI 2015: IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems. San Diego State, California, USA. http://mfi2015.sdsu.edu/

14–18 September
Summer School on Experimental Methodology, Performance Evaluation and Benchmarking in Robotics. Intur Bonaire Hotel, Benicassim, Castellon, Spain.

28 September–2 October
IROS 2015: IEEE/RSJ International Conference on Intelligent Robots and Systems. Hamburg, Germany. http://www.iros2015.org/

Digital Object Identifier 10.1109/MRA.2015.2465892
Date of publication: 11 September 2015

13–16 October
ICCAS 2015: International Conference on Control, Automation, and Systems. BEXCO, Busan, Korea. http://2015.iccas.org/

18–20 October
SSRR 2015: IEEE International Symposium on Safety, Security, and Rescue Robotics. Purdue University, Indiana, USA. https://robotics.purdue.edu/SSRR2015/index.html

3–5 November
Humanoids 2015. Seoul, South Korea. Call for Papers Deadline: 30 June 2015. http://www.humanoids2015.org/main/

6–9 December
ROBIO 2015: IEEE International Conference on Robotics and Biomimetics. Zhuhai, China. Call for Papers Deadline: 25 July 2015. http://ieee-robio.org/2015/

12–13 December
SII 2015: IEEE/SICE International Symposium on System Integration. Nagoya, Japan. Call for Papers Deadline: 31 August 2015. http://www.si-sice.org/SII2015/

2016

16–21 May
ICRA 2016: IEEE International Conference on Robotics and Automation. Stockholm, Sweden. Call for Papers Deadline: 31 August 2015. http://www.icra2016.org/

8–11 July
AIM 2016: IEEE/ASME International Conference on Advanced Intelligent Mechatronics. Ottawa, Ontario, Canada.

Turning Point (continued from p. 192)

specific scientific subject. For instance, for mathematical sciences, ensuring reproducibility is quite straightforward since it requires clear reasoning and a complete proof of the presented results, which can be (easily) followed by the readers. Once this is done, the results become part of the common body of knowledge for the community and do not need to be proven again.

The concept is usually specialized and varies for each particular research community.

Conversely, for empirical sciences (such as life sciences, social sciences, and, in the area of interest of the IEEE, e.g., devices/circuits/systems implementation and characterization), the quest for RR involves, at a minimum,

1) a clear description of the methodology followed in a particular study or experiment
2) a detailed explanation of the laboratory procedures/protocols used
3) a thorough statistical analysis of the results obtained, highlighting their significance
4) the complete sharing of the data associated with the study/experiment
5) the sharing of the code and the features of the run-time environment that has (possibly) been used to produce the data.

The extent of this list clearly shows the intrinsic difficulties in guaranteeing reproducibility in this context.

For the computational sciences, which are a subject of interest for several scientific communities within the IEEE, guaranteeing RR mainly involves points 1), 4), and 5), thus resulting in an intermediate difficulty level.

Q: Why is RR important?

GS: Since the times of Galileo and Boyle, the basis of science has been the capability to replicate the results produced by other researchers, to build on their discoveries, to advance knowledge and technology. In other words, reproducing previous results to show the advantages of the proposed innovative methodologies or techniques has always been the key to progress in science. Using Isaac Newton's famous expression, one can summarize this concept: "If I have seen further, it is by standing on the shoulders of giants" [3].

While this scientific approach worked remarkably well for centuries because of the ability of the scientific community to discover and correct mistakes and refine or completely change flawed theories and erroneous methodologies, in recent years something seems to have gone wrong in the self-correcting mechanism of science, particularly in the area of life sciences. Even if one does not consider the most outrageous cases of fraudulent research, such as the famous stem-cell scandal (which involved the retraction of two papers published in 2004 and 2005 in *Science* [4]), several recent studies have highlighted the impossibility of reproducing the results published in the vast majority of the papers under investigation. More precisely, and by way of example, according to Begley and Ellis [5], only 11% of 53 studies in the area of preclinical cancer drugs were reproducible, while Ioannidis et al. [6] show that this was also true for two of 18 papers in bioinformatics. What is worse is that similar findings have made their way into the general public press [7] and generated in the public opinion an increased sense of unease with respect to the way in which science operates.

A systematic adoption of RR practices is certainly necessary to reverse this worrisome trend. At the same time, its implications are far more important than this. RR is, in fact, fundamental since the following hold.

- It will foster growth in the capabilities for collaboration among scientists, which will help to overcome the increasing challenges posed by the rising number of multidisciplinary collaborations.
- It will produce an increase in the rate of innovation: researchers will advance technology more easily, and practitioners will develop new products faster.

This is, of course, a future that every scientist and practitioner will welcome as important steps forward for humanity.

Q: What is RR for the IEEE, and why it is important?

GS: IEEE is first and foremost a professional organization, and its publishing enterprise exists as a service to the community. One of the reasons reproducibility is important for the institute is that there are more indications that RR may actually soon be incentivized (if not mandated) by funding agencies in a similar way with respect to what has happened in recent years for open access. Another reason underlying its importance is that RR may simply become a more pressing request by the IEEE members and authors. More scientists are, in fact, interested in increasing the visibility of their discoveries: preliminary studies show a greater impact for those scientific works that share supplemental material together with the paper itself [8].

Consequently, simply because (part of) the IEEE community will need it, the development of an infrastructure supporting RR (at least in terms of storing/reusing data, code, and algorithms) may become, in my view, a pressing need for the IEEE in the not so distant future.

There are, however, other advantages that the adoption of RR will offer to IEEE. First, as previously mentioned, because of RR, the information made available through the IEEE conferences and publications will be more visible and directly usable by both scientists and practicing engineers. Furthermore, the adoption of RR will help the readers to navigate the large quantity of papers available on a specific subject. By straightforwardly reproducing results, readers will directly test the advantage of a technique with respect to a different one. Finally, promoting RR will make it easier to discover possible false (or inaccurate) results and help the IEEE to maintain its reputation as a world-class scientific/professional organization.

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

Q: This special issue on reproducibility and measurability of robotics research demonstrated a high interest from the community of the IEEE Robotics and Automation Society (RAS). What opportunities do you foresee for linking this interest to future initiatives in this area which could be launched at the IEEE level?

GS: Developing the necessary infrastructure for supporting RR as well as the best practices associated with it (e.g., in terms of the review process of the data, code, and algorithms associated with the paper) will require substantial work and support from many different IEEE communities. The RAS has already made significant steps in these directions and its experience will be truly precious for the entire organization.

Q: Do you consider research in the robotics and automation domains a key area for applying the principles, methods, and tools for aiming at RR?

GS: Absolutely. Robotics and automation is one of the best areas to apply and test any best practices that the IEEE will develop in terms of RR. In fact, the robotics and automation domains rely on mathematical science as well as computational and experimental ones, so that these experiences pertain to all kinds of reproducibility mentioned in the answer to the second question.

Q: Do you consider these topics important for the training of a new generation of researchers in engineering?

GS: I consider them fundamental. Adopting RR will, in fact, truly change the culture and will require substantial additional effort from the authors publishing with the IEEE. This is, of course, a process that cannot be enforced but only reinforced. We need, therefore, to educate the community, especially the young professionals, to comprehend and embrace the benefits that RR can bring from all different perspectives: authors (visibility increase), users (enhancement in the exploitability of results, increase in capability of recognizing fundamental results) and humanity as a whole (increased rate of innovation).

References

[1] Wikipedia. Reproducibility. [Online]. Available: http://en.wikipedia.org/wiki/reproducibility

[2] J. B. Buckheit and D. L. Donoho, "WaveLab and reproducible research," Dept. of Statistics, Stanford Univ., Stanford, CA, Tech. Rep. 474, 1995.

[3] H. W. Turnbull, Ed., *The Correspondence of Isaac Newton: 1661–1675*, vol. 1. London, U.K.: Royal Society at Univ. Press, 1959, p. 416.

[4] Special online collection: Hwang et al. controversy—Committee report, response, and background. *Science*. [Online]. Available: http://www.sciencemag.org/site/feature/misc/webfeat/hwang2005/

[5] C. G. Begley and L. M. Ellis. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*. [Online]. *483(7391)*, pp. 531–533. Available: http://www.nature.com/nature/journal/v483/n7391/full/483531a.html

[6] J. P. A. Ioannidis et al. (2009). Repeatability of published microarray gene expression analyses. *Nature Genet*. [Online]. *41(2)*, pp. 149–155. Available: http://www.nature.com/ng/journal/v41/n2/full/ng.295.html

[7] How science goes wrong. (2013, 19 Oct.). *The Economist*. [Online]. Available: http://www.economist.com/printedition/2013-10-19

[8] H. A. Piwowar, R. S. Day, and D. B. Fridsma, "Sharing detailed research data is associated with increased citation rate," *PLoS One*, vol. 2, no. 3, e308, 2007, DOI: 10.1371/journal.pone.0000308.

From the Field *(continued from p. 188)*

Importantly, the system also allows researchers to create their own specialized, operational tasks to run in natural settings. We claim that this method can be applied to all kinds of applications.

An opportunity to try this out arose in July 2014 at the Arkaroola Mars Robot Challenge organized by the Mars Society Australia. Four student teams brought six field robots to a test site in Arkaroola, a remote desert station in South Australia. The machines embodied the students' design concepts for assistant robots for astronauts performing tasks on the Martian surface (Figure 1). A selection of six standard DHC–NIST–ASTM benchmarks, together with three operational tests specific to surface operations in harsh Mars-like terrain, was conducted over 12 days. For example, we had the robots search a gullied slope for a hidden target object, which had to be photographed, collected, and returned to the operator.

The test details and results will be formally presented in September at Towards Autonomous Robotic Systems 2015 in Liverpool, United Kingdom, but in brief, we found that most, but not all, tests worked well, provided one practices the procedures and allows enough time (Figure 2). We were not only able to gather a good deal of standard performance data, but we were able to use it later to make real design improvements to two of the robots. Our test program could accommodate unmanned aerial vehicles as well as ground machines: one participant was able to score highly on many tasks using a small quadrotor, suggesting very high utility of a (suitably modified) drone for future Mars explorers.