The Role of Experiments in Robotics Research

Fabio P. Bonsignorio, John Hallam, Angel P. del Pobil and Raj Madhavan

Abstract— **Experiments are essential ingredients of science, both to confirm/refute a theory and to discover new theories.**

It is a common belief that experimentation in robotics has not yet reached a level of maturity comparable with that reached in traditional science. In this workshop, we will discuss fundamental issues about the role of experiments in robotics, such as how can results be replicable and refutable on the one hand, and quantitatively comparable according to community-endorsed metrics to enable a faster cumulative progress, or even appreciate disruptive changes.

These issues, when viewed in the context of some general principles about experiments in science and engineering, will lead us to derive some insightful considerations on the role of experiments in robotics. A key point to allow replication and comparison of results is having adequate data support: all the data necessary to repeat a given experiment, how to achieve it with today's digital media will be addressed.

I. Introduction

Most forms of robotic system performance measurement, evaluation, comparison, etc. involve practical experimentation, which must be carried out responsibly and reported well. Recently, the interest in experimental methodologies increased dramatically within the robotics community, both from researchers and from funding agencies, according to the idea that good experimental activities could reduce the gap between research and industrial applications.

Some projects have been funded by the European Commission and a series of workshops (see below) have been held in the latest years.

Despite the ongoing efforts and the recognized importance of experiments for rigorously evaluating new approaches and for reporting them in an objective and complete manner, these ideas have not yet become really part of current practice, as it can be seen by having a look at recently published papers, see [18]. This may be due to the difficulty to perform time-consuming experiments and to the weak awareness of experiments as fundamental elements in the development of a robotic system.

The aim is to discuss fundamental issues about the role of experiments in robotics, and namely as already noticed above, issues such as how can results be replicable and refutable on the one hand, and quantitatively comparable according to community-endorsed metrics, on the other hand. As both replication and benchmarking are needed to foster a cumulative advancement of robotics, and even to correctly appreciate disruptive innovation in the science and technology of robots. The recent evidence of community interest suggests that it is willing to take steps to improve in this area.

A deep discussion of these issues, when viewed in the context of some general principles about experiments in science and engineering, will lead us to derive some insightful considerations on the role of experiments in robotics.

The ICRA 2010 workshop on the role of experiments in robotics is a joint initiative of the

Fabio P. Bonsignorio is CEO and founder of Heron Robots srl V.R.Ceccardi 1/18, 16121 Genova, Italy and is Santander Chair of Excellence in Robotics at UC3M in Madrid. (phone: +39-339-8406011; fax: +39-0183-730265; e-mail: fabio.bonsignorio@heronrobots.com)

John Hallam is with the Mærsk Mc-Kinney Møller Institute, University of Southern Denmark (SDU), Odense, Denmark (e-mail: john@mmmi.sdu.dk).

Angel P. del Pobil is with Jaume I University (Spain), and founder director of the Robotic Intelligence Laboratory , Castelló de la Plana, Spain (e-mail: pobil@icc.uji.es).

Raj Madhavan is an assistant research scientist with the Institute for Systems Research (ISR) and a guest researcher with the Intelligent Systems Division of the Manufacturing Engineering Laboratory at the National Institute of Standards and Technology (NIST), 100 Bureau Dr. Stop 8230, Gaithersburg, MD 20899, U.S.A. (phone: +1-301-975-2865; e-mail: raj.madhavan@nist.gov)

new IEEE-TC on Performance Evaluation and Benchmarking of Robotic and Automation Systems (PEBRAS) and the EURON Special Interest Group on Good Experimental Methodology for Robotics (SIG GEM). The proposers are co-chairs of IEEE-TC PEBRAS and/or EURON SIG GEM and have jointly co-organized over 10 successful related events in the last 5 years, such as 4-in-a-row workshops on Benchmarking and Performance metrics at IROS (2006-2009), two workshops at RSS (2008-2009) and the Performance Metrics for Intelligent Systems Workshop (PERMIS) series, which began in 2000.

One of the core activities of Euron, the European Union funded network of excellence in robotics was devoted to the definition and diffusion of proper benchmarking procedures for intelligent and robotic system. A few years ago, EURON started a Special Interest Group on Good Experimental Methodology and Benchmarking in order to address these issues. Since then many discussion has been carried on, see [7,5].

A major output of the Euron SIG GEM is a set of guidelines for good experimental and reporting practice.

The GEM guidelines, [17,29], don' t apply to every kind of papers: 'position papers', summarizing a point of view about a general issue, like this one, 'concept' paper describing at concept level a research to come, papers reporting on field testing of a robotic solution, papers usually interesting, informative and worth publishing, are out of the scope of these guidelines. We pay special attention to the papers we define as 'experimental papers': the papers where a theoretical claim is based on simulation or field experiments of a technological approach, an algorithm or a set of algorithms.

A rational research enterprise should not start from scratch every time, but should build on the results already obtained by other people, groups, or organizations. It can be easily verified that more than 90% of published robotics research papers include some kind of field or simulation testing. Despite that they are often not comparable to similar ones and pose serious issues to be reproduced.

These papers are usually interesting, the described testing activities have in many cases a 'rhetoric' purpose, but not very often can be considered 'experimental' papers with the meaning given above.

According to the commonly received idea of the scientific method, it is easy to agree on the fact that, if we are following a scientific methodology, when we face an experimental paper we should be able to:

- validate the results by replicating them
- compare the results in term of the chosen performance criteria.

A common objection to these points is that robotics research looks at quite diversified problems and objects so that comparisons are in principle or in practice impossible.

It is useful, under this respect, to compare with older and more established research domains, like medicine and life sciences, where complexity and variety of the studied objects is not minor than in robotics.

In any case if we aim to do 'scientific' claims on which technical approaches are 'better' according to a set of given criteria for a given application set of tasks and environments, a kind of experimental methodology is needed. And applies whether we are in a cumulative phase in the development of our discipline or in presence of a 'disruptive' creative paradigm shift, as somebody is claiming, especially in the cognitive science community.

II. Epistemological Issues

As observed above it is not always easy to verify if and by which measure new procedures and algorithms proposed in research papers constitute a real advancement and can be used in new applications.

As it is not clear which is the state of the art from both a quantitative and quantitative standpoint it is not impossible that new more successful implementations of concepts already presented in literature, but not yet implemented with exhaustive experimental methodology, are ignored. Since benchmarking procedures, allowing to compare the actual practical results with reference to standard accepted procedures, are not widely accepted. It is also difficult to rationally compare different paradigms like for instance fully actuated or passive walkers, or top-down symbolic planning against self-organized behaviors.

Robotics concepts and methods are used in different contexts for different purposes. Robotics is 'science' when it deals with reverse engineering of animal locomotion or investigate how a natural system can exhibit cognitive or autonomy capabilities, and it is 'engineering' when it develop a new system to cope with a human need. If we want to demarcate robotics from astrology a 'scientific methodology' is needed in both cases. It is also possible to see robotics as the science of intelligent physical agents ('embodied cognition'). Replication/reproducibility of experiments and quantitative performance measurement procedures are needed to define robotics research as a scientific enterprise. The word 'experiment' itself is not widely used in our field. What we should define as an 'experiment' in robotics? Which meaning should we give to 'replicability' in our context?

Although, it is known K. Popper defined in a very strict way the requisites for a discipline to be considered 'scientific' focusing mainly on physics, in other disciplines, in social science, management and economics exact repetition is often seen as a limit case. Only when the model fails clearly in a number of variated experimental setup it is considered 'not replicable'. Nevertheless, as already noticed, all disciplines aiming to be considered 'scientific' a concept of experiment replication/reproduction and more generally a concept of 'verification' of theory through experiments, [24,25,26,27,28].

It seems unlikely to successfully import into robotics a too strict verification concept: as already noticed, the huge variability of robot machines, tasks, and application environments limits the replication and comparison of results.

Interesting hints come from the epistemological analysis of biology which seems to share some of the 'foundational issues' affecting robotics. In comparison to other scientific fields, like, namely, physics, the status of biology as a scientific discipline requires an extension of the usual methodological concepts as they are commonly received.

The definition of what should be considered a 'law of nature' in biology raises a number of issues. For reasons not very different from those raised from robotics research. The laws are usually not universal but apply to specific species: the Mendel laws apply to species with sexual reproduction, but not to all living species.

Almost every theoretical enunciate refers to a species or a set of species and has stochastic characteristics.

Systems are usually very complex, involve a huge numbers of variables and work in open ended stochastic environments. The same function, for example flight, can be performed in many different ways. The wing morphology and dynamics of a fly are quite different from those of a bird. On an other end, the wing of the penguins are used to stabilize swimming.

An interesting point is that the laws regarding a specific function in a species become true at a specific time, as a new function evolve, as depicted in fig. 1., and only if some initial conditions occur.

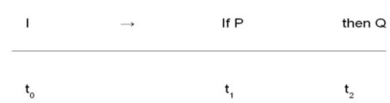


Fig1. Time dependence of biological laws

In other cases the high level behaviors of a system emerge from the superposition of many non linear underlying processes, see fig. 2, for example at neural or biomolecular level.

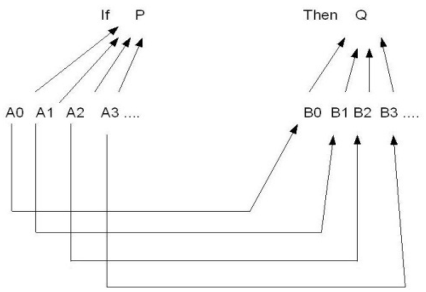


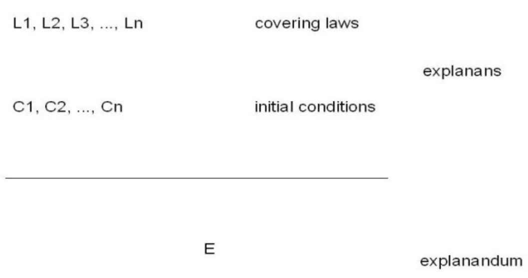
Fig.2 'Causality at different levels'.

Despite that it is hard to deny a scientific status to biology. This led also in the recent past to a rich epistemological discussion on the 'scientific' nature of biology. S. Mitchell proposed an interesting pragmatical classification scheme for scientific disciplines, [33].

According to her view, 'laws of natures', can be classified in a continuum in a three axis volume in term of abstraction, (deterministic) strength, stability (in time).

Interesting analysis of the issues were provided by Schlick ('Schlick's problem'), [37], pointing out that a given set of data can be interpreted 'ex-post' in many different logically consistent ways and more recently by Nagel, [32], and Goodman, [34].

A possible approach, [31], is to formally define a 'Question' Q as a triple (P_k, X, R). Where P_k is the question theme, X = [P_1,...P_k,...] is the contrast class and R is a relevance relation. A is a valid answer iff P_k in X is true, A is true, R: (P_k, X)-> A.



.

Fig.3 Hempel-Oppenheim Schema

In the conceptual schema represented in Figure 3, which summarizes the Hempel-Oppenheim model of scientific knowledge,[35,36], all the logical enunciate have a probabilistic truth value. All the examples provided above seem to be general enough to ground a scientific methodology for robotics.

We need a precise and complete list of laws invoked for the explanation, a precise and complete list of initial condition (system hw/sw architectures, environments, tasks), a precise definition of what is 'explained' or proven. And we must accept the fact as we operate in open ended stochastic environments our theoretical claims, 'enunciate', have to be of probabilistic nature.

III. Discussion

Robotics research spans over a wide variety of applications domain and tasks with very different morphologies, dynamics and control approaches and implementations. This makes its investigation domain similar to that of biology.

In both cases the results are highly context dependent.

Following the 'contextual' approach proposed above the triple identifying a 'question' above must be defined every time. For instance, if we think to locomotion it will be different to have legs or wheels. If you talk about SLAM it is different to talk about laser scans or visual information.

This should not be seen as preventing the adoption of a rigorous scientific methodology. Examples of 'context dependent' laws of nature date back to very beginning of modern physics: Kepler's laws were initially introduced for the solar system and as such they are not 'general' at all.

The theoretical claims ('enunciate') we are going to support will probably be restricted to laser scan based SLAM wheeled mobile robot settings (and sometimes it is probably necessary to be even narrower in scope).

How can we enable the replication of research results?

Starting from Figure 3 conceptual schema we can define a number of requisites for a robotic

experiment, see [17,29].
We define an experiment a real world verification by means of a set of empirical tests of an answer to a significant engineering or scientific question about a robotic (or robotics-related) system.
In general experiments may be conducted using simulation as a tool although this prevents a straightforward generalization to the real world. The advantage is that simulation allows much more variations in the parameter space.
The 'context' of the investigation must be clearly defined in advance: system structure, control methods, task sets, environment sets.
An experimental paper should address an interesting engineering (or scientific) question, in the sense recalled above. Such questions will generally refer to quantitative relationships between system, task or environment parameters and some system performance metrics. The performance metrics being studied must be explicitly motivated and clear hypotheses must be done about the parameters of the system(s), environment and tasks.
The criteria for evaluating results should be stated and, where necessary, justified. The performance criteria being studied must be measurable.
All data regarding the design, the system parameters, the environment and tasks should be provided in order to make possible to reproduce the work.
The benchmarking criteria must be explicitly expressed: this means that the metrics and their measurement operational procedures in a wide enough set of initial condition must be given.
The statistics distributions of the system, task and environment characteristics parameters must be given.
The 'adverse' events, those possibly contradicting the major claims, must be reported. The anomalies have to be justified in the context of the maintained theoretical enunciates.

IV. EXAMPLE

As an example we discuss the requisites of a replicable robotics experiment in visual servoing, according to [17]

Visual servoing control the movement of robot (video assisted mobile robots or manipulators) on the basis of feedback coming from a video device, like a video camera. This robotics sub-field is, as it usually happens in robotics, an interdisciplinary domain integrating computer vision, robotics, kinematics, dynamics, control and real-time and embedded systems. Papers usually present a theoretical part describing control methods, visual features and models and show results obtained by field experiments or by simulation.
This example is relevant because formal proof are very difficult if not impossible in many if not most cases, as a consequence experimental work is necessary to assess the potential of different approaches to control. When dealing with this topic theoretical 'enunciates' must in many cases be based on experimental proofs.
Here below we list a number of requisite necessary for experiment replication (and performance comparison).

Assumptions

For a visual servoing systems there typical which must be detailed. A non exhaustive list is given here:

- the visual features
- scene 3D model
- the kinematics model of the robot.
- dynamics model of the robot.

Plus the list related to image processing:

- background characteristics (homogeneous or if not color and luminance distributions)
- lighting conditions
- robustness to outliers in feature detection
- others inherent to real life experimentation.

Performance criteria

Generally speaking these criteria measure the convergence of the system to a predefined goal.
Non exhaustive list:

- the time of convergence
- the trajectories of the visual features in the image plane

- the 3D trajectory of the robot
- computation time
- positioning error after convergence.

A special attention must be paid to stability and robustness against image noise, the errors in the models (object, camera, robot), and the control parameters.

Measured characteristics

An unequivocal procedure to derive the quantitative aspects of the system must be given. For example visual features can be directly obtained from the video camera.
For manipulators what is directly measurable are the generalized joint angles while the end effector 3D trajectory must be estimated by the (direct) kinematic model.
Calibration procedures for the robot relevant characteristics and camera must be described.
In experiments the visual features (at least) must be variated and the variation policy documented.

Implementation Information

The information given above don't allow by themselves the replication of results.
There more data needs than in other kind of papers:

- Visual servoing system configuration environment (either real or simulation) should be described in detail: in-hand vs. external camera, etc.
- model and control parameters
- Ground truth for robot positioning and the environment
- Technical specification of the hardware platform
- Technical specifications of the camera (model, frame rate, resolution, etc.).
- Computer specifications (at least, processor and amount of memory, o.s., relevant configuration details)
- sw libraries (they should be available at least as linkable components) list and configuration

Probably the adoption of widely known sw libraries like ViSP, VXL, OpenCV may ease replication.

Parameter and variable distribution

Statistical distributions of all relevant parameter must be given (as in an open ended stochastic environment results will have a probabilistic formulation). This is by the way quite common in clinical research.

Detailed list of findings

The list of findings in the discussion/conclusion section should be against a detailed list of criteria within a detailed list of conditions as recalled above
For example better convergence speed, robustness /weakness against certain parameters, behavior with respect to current technology visual servoing systems:

- visual features moving of the field of view
- workspace and singularity issues

The findings listed in a paper might be negative: the given algorithm in our test conditions fail under the listed set of conditions with respect to the listed series of criteria.

V. Conclusions and Future Work

Robotics research deals, at least, with two different set of challenges: the reverse engineering of intelligent systems with can observe in nature and the development of new 'intelligent/cognitive' machines to cope with human needs. No doubts the first must be regarded as a scientific domain: to a certain extent it might be seen as a biology sub field. The second one define robotics as an engineering discipline. Even the second set of objectives requires to define an appropriate scientific methodology as 'modern engineering' is characterized by scientific methodology. It is thought that in both these situations the epistemological model based on 'context' discussed above for biology and extended to robotics may provide a working framework.

There are implication of the reporting process.

We may think of theoretical/concept papers, proof of concept papers, and experimental papers , as we have started to define here, as steps in a research idea 'life-cycle'. We believe that more paper of the 'experimental' kind would greatly help the research activities in robotics and the industrial exploitation of the results.

Moreover we envision a new kind of replicable/reproducible 'papers'.

A 'replicable' report should (at least?) include:

1) 'description' : a journal paper text+figures+ multimedia ….according to GEM Guidelines (or similar)

2) Data sets (similar to IJRR 'Data paper')

3) Complete 'code' identifiers and or downloadable code (executables may be enough)

4) 'HW' description or HW identifier (if it is identifiable)

To fill all this information is impossible in a journal printout, but it is quite easy in a web 2.0 publishing facility.

We believe that in the future we should deploy such kind of reporting processes, going beyond the traditional journal 'format'.

ACKNOWLEDGMENTS

This paper is largely inspired by the discussions and the contributions within the GEM SIG, the benchmarking activities of Euron Noe, and the Permis series conferences. Among the people involved in the discussions: Francesco Amigoni, Elena Messina, Nicola Basilico, Diego Alonso C´aceres, Daniele Calisi, Enric Cervera, Bridget Hallam, Lino Marquez, Matteo Matteucci, Javier Minguez, Jos´e Neira, Francisco Ortiz, Enrico Pagello, Mario Prats, Monica Reggiani, Domenico Sorrenti, Kasper Støy, Juan Tardos, Vittorio Ziparo, Richard Vaughan, Herman Bruyininkx and many others attending the SIG meetings, the benchmarking workshops and the Permis series conferences.

A special thanks to Antonio Bicchi, Tim Smithers and others who provided challenging comments. And to John Albus from NIST who pioneered the discussion on the measure of intelligent system performances.

REFERENCES

[1] J. S. Albus, "Metrics and Performance Measures for Intelligent Unmanned Ground Vehicles". In Proceeding of the performance Metrics for Intelligent System Workshop, 2002.

[2] A. P. del Pôbil, "Why do We Need Benchmarks in Robotics Research?", International Conference on Intelligent Robot and Systems, Beijing, China, 2006.

[3] http://www.isd.mel.nist.gov/PerMIS_2009/index.htm

[4] http://www.robot-standards.eu/

[5] http://www.robot.uji.es/benchmarks/index.html

[6] http://www.rawseeds.org/

[7] http://www.herronrobots.com/EuronGEMSig

[8] http://www.isd.mel.nist.gov/projects/autonomy_levels/

[9] http://www.robocup.org

[10] http://www.darpa.mil/grandchallenge/

[11] J. W. Crandall and M. A. Goodrich, "Measuring the Intelligence of a Robot and its Interface". In Proceeding of the performance Metrics for Intelligent System Workshop, 2003.

[12] J. Minguez, J. Osuna, and L. Montano. "A 'Divide and Conquer' Strategy based on Situations to achieve Reactive Collision Avoidance in Troublesome Scenarios". In ICRA, New Orleans, USA, 2004.

[13] L. Olsson, C.L Nehaiv and D. Polani, "Information Trade-Offs and the Evolution of Sensory Layouts", In *Proc. Artificial Life IX*, 2004.

[14] M.Lungarella and O. Sporns, "Mapping Information Flow in Sensorimotor Networks", *PLOS Computational Biology*, 2, 10, pp. 1301-1312 , 2006.

[15] A.Lampe, R.Chatila, "Performance measures for the evaluation of mobile robot autonomy" , IEEE International Conference on Robotics and Automation (ICRA'06), Orlando (USA), 2006.

[16] J.Hallam and G.Hayes, "Benchmarks for mobile robotics?" In Towards Intelligent Mobile Robots: scientific methods in mobile robotics, Manchester University, School of Computer Science, 1997.

[17] F. Bonsignorio, J. Hallam, and A. P. del Pobil, "Good Experimental Methodology - GEM Guidelines," http://www.heronrobots.com/EuronGEMSig/Downloads/GemSigGuidelinesBeta.pdf, 2007.

[18] F. Amigoni, M. Reggiani, and V. Schiaffonati., "An insightful comparison between experiments in mobile robotics and in science," Autonomous Robots, vol. 27, no. 4, pp. 313–325, 2009.

[19] R. Madhavan, C. Scrapper, and A. Kleiner, "Special issue on characterizing mobile robot localization and mapping," Autonomous Robots, vol. 27, no. 4, pp. 309–481, 2009

[20] S. Hartmann, "The world as a process: Simulations in the natural and social sciences," in Simulation and Modeling in the Social Sciences from the Philosophy of Science Point of View, Hegselmann, R. et al., Ed. Kluwer, 1996, pp. 77–100

[21] https://www.ctnbestpractices.org/

[22] http://ctep.cancer.gov/handbook/hndbk_7.html

[23] http://www.cancer.gov/

[24] J. Pfeffer, "Barriers to the advance of organizational science: Paradigm development as a dependent variable". Academy of Management Review, 18: 599–620, 1993.

[25] K. Popper, *The logic of scientific discovery*.,Hutchison, London, 1959

[26] T. Kuhn, *The structure of scientific revolutions* (2nd ed.), University of Chicago Press, Chicago, 1970.

[27] I.Lakatos, "Criticism and the Methodology of Scientific Research Programmes",in Proceedings of the Aristotelian Society, vol. 69, pp. 149-186, 1968.
[28] P.K.Feyerabend, *Against Method*, Verso, London, 1975.
[29] F. Bonsignorio, J.Hallam and Angel P. Del Pobil (eds), GEM Guidelines, Euron GEM Sig Report, (2008)
[30] C. Ruhla, The Physics of Chance, Oxford University Press, Oxford, 1992.
[31] G. Boniolo, A Contextualized Approach to Biological Explanation, Philosophy, 80, 219-247, 2005.
[32] E. Nagel, The Structure of Science. Problems in the logic of the scientific explanation, Harcourt, New York, 1961.
[33] S.D.Mitchell, Dimensions of Scientific law, Philosophy of Science, 67, 242-265, 1997.
[34] N.Goodman, Fact, Fiction and Forecast, Harvard University Press , Cambridge (MA, USA), 1954.
[35] C.G. Hempel, Aspects of Scientific Explanation and other essays in the Philosophy of Science, Free Press, New York, 1965.
[36] C.G. Hempel, P. Oppenheim, Studies in the logic of explanation, Philosophy of science, 15, 1948.
[37] M. Schlick, Die causalitaet in den gegerwaertigen physik, Die naturwissenschaften, 19, 145-162, 1931.