Defining the Requisites of a Replicable Robotics Experiment

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

***Abstract*— There is, in the robotics community, a growing awareness of the difficulty to compare in a rigorous quantitative way the many research results obtained in the many different application areas of the field. It is thought that, if we aim to consider robotics a real 'science' or a branch of modern engineering, we must pay some attention to the experimental methodology. In this paper we focus on the issues raised by the replication/reproducibility of results, which under one respect are a cornerstone of any scientific methodology, and on a different respect are a basic pre-requisite to compare different methods for common problems proposed in the literature.**

I. Introduction

The number of published papers at conferences and in journals is steadily increasing and the number of journals and conferences are, too. For any robotics application different methods, sometimes based on remarkably different principles, are proposed. Generally these methods, algorithms, procedures are implemented on different, sometimes very different, hw/sw architectures and environments.

This create many opportunities for the industrial exploitation of results. Nevertheless in many cases it is not easy to compare the relative strength of the methods proposed for the same functionality, while in many, if not most, cases it is extremely time consuming and sometimes impossible even to simply replicate the published experiments.

This is a bottleneck for the industrial exploitation of results as it is difficult to assess the potential of a newly proposed technological approach and even to assess objectively the state of the art in any specific subfield. This also slows 'basic' robotic research itself, as, in general, there is a limited cross exploitation of results between the different labs.

A couple of years ago, EURON, the European robotics network, started a Special Interest Group on Good Experimental Methodology and Benchmarking in order to address these issues. Since then many discussion ha been carried on, see [7,5], and some starting guidelines have been drawn.

The GEM guidelines, [25], 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, 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

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).

considered 'experimental' papers with the meaning given above.

There is a broad agreement on the fact that 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 very diversified problems and objects so that comparisons are in principle impossible.

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

In any case if we aim to do 'scientific' claims on which technical approaches are 'better' according to given criteria for a given application set of tasks and environments a kind of experimental methodology is needed in order to be able to ground the advancement of research on a shared quantitative language. And this is true 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 qualitative 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 that 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, [15,16,17,18,19].

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 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 refer to a species or a set of species and has stochastic characteristics.

Systems are usually very complex, involve a huge numbers of variable 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 a penguin 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.

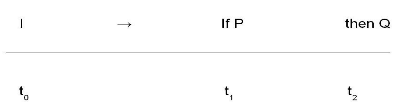

Fig1. Time dependence of biological laws

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.

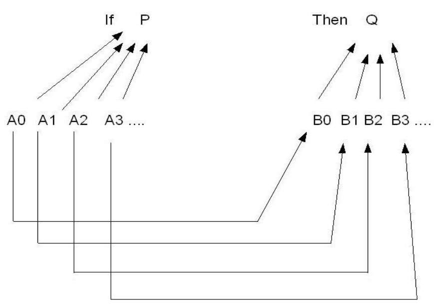

Fig.2 'Causality at different levels'.

Despite that it is hard to deny a scientific status to biology. This led to a rich epistemological discussion on the 'scientific' nature of biology. S. Mitchell proposed an interesting pragmatical classification scheme for scientific disciplines, [29].
According to her view, 'laws of nature', 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'), [33], pointing out that a given set of data can be interpreted 'ex-post' in many different logically consistent ways and more recently by Nagel, [28], and Goodman, [30].

A possible approach, [27], is to formally define a 'Question' Q as a triple (Pk, X, R). Where Pk is the question theme, X = [P1,...Pk,...] is the contrast class and R is a relevance relation. A is a valid answer iff Pk in X is true, A is true, R: (Pk, X)-> A.
.

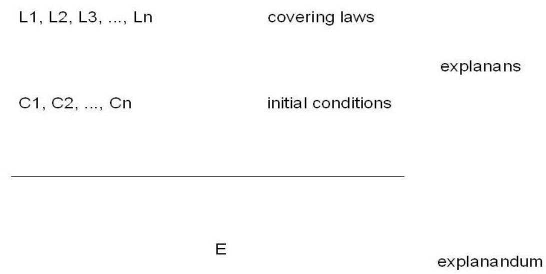

Fig.3 Hempel-Oppenheim Schema

In the conceptual schema represented in Figure 3, which summarizes the Hempel-Oppenheim model of scientific knowledge,[31,32], all the logical enunciate have a probabilistic truth value. All the example provide above seems 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 span a wide variety of applications domain and tasks with very different morphologies, dynamics and control approaches and implementations. This make its investigation domain similar to that of biology.

In both case 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 [25].

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 allow 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.

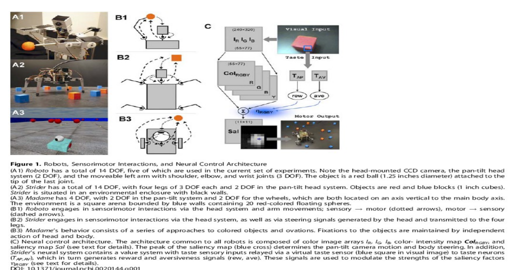

Fig.4. Sensory motor coordination information metrics examples (Lungarella and Sporns, 2006)

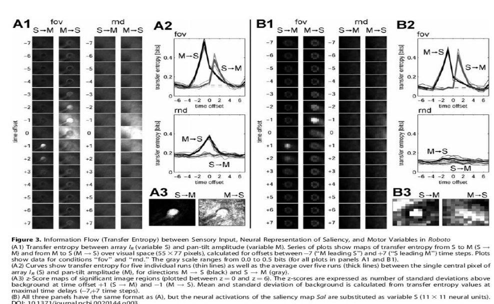

Fig.5. Sensory motor coordination information metrics measures (Lungarella and Sporns, 2006)

A serious issue is given by the open ended environments where robot systems are supposed to operate. The environment and tasks, to achieve enough flexibility, require to be characterized from a stochastic perspective.

When 'cognitive' capabilities are involved it is

necessary to compare the 'complexity' of different stochastic and time varying set of tasks and environments.
It is likely that these metrics can be derived from information metrics related to Shannon entropy.
It is worth notice as in [13] and [14], see fig. 4 and 5, from [14], and in other experimental works ‘entropy measures’ on the ‘sensory-motor’ coordination of different ‘robotics’ equipment have shown that information metrics can be used to classify, at least, and to get an insight on (semi) autonomous robotics devices, which show an ‘emergent behavior’, while, in [15], entropy measures are used to rank environment complexity, with reference to the navigation task, see fig. 3.
In [12] an approach integrating task and environment complexities is proposed.
Approaches like those quoted above are of wide use in computational neurosciences.
HRI (Human Robot Interaction) experimental research is sometime conducted by means of protocols deriving from psychology, while there are comparative studies of animal and robot group dynamics.

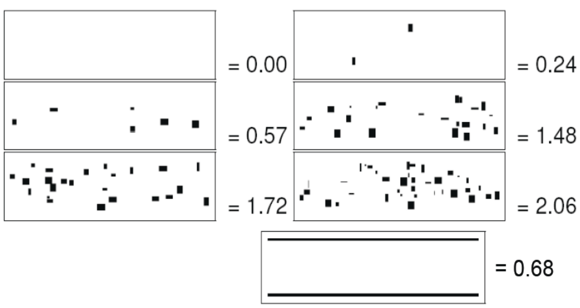

Fig. 6. Environment complexity measures (Lampe, Chatila, 2006)

Possibly to allow a complete quantitative characterization of robot behaviors, we need a kind of unification between fields so far considered separated like control and information theory, general AI and system dynamics.

IV. Example

As an example we discuss the requisites of a replicable robotics experiment in visual servoing, according to [25].
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 requisites necessary for experiment replication (and performance comparison).

A. Assumptions

For a visual servoing systems there are typical aspects 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.

B. 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.

C. 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.

D.Implementation Information

The information given above don't allow by themselves the replication of results.
There are 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.

E.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.

F.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 we 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 defines 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.

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.

Acknowledgments

This paper is largely inspired by the discussions and the contributions within the GEM SIG, with many people, among them: Francesco Amigoni, Nicola Basilico, Diego Alonso C´aceres, Daniele

Calisi, Enric Cervera, Bridget Hallam, Lino Marques, Matteo Matteucci, Javier Minguez, Jos ´e Neira, Francisco Ortiz, Enrico Pagello, Mario Prats, Monica Reggiani, Domenico Sorrenti, Kasper Støy, Juan Tardos, Vittorio Ziparo, and many others attending the SIG meetings.

A special thank to Antonio Bicchi, Tim Smithers and others who provided challenging comments.

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_2007/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]https://www.ctnbestpractices.org/

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

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

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

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

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

[23]I.Lakatos, "Criticism and the Methodology of Scientific Research Programmes",in Proceedings of the Aristotelian Society, vol. 69, pp. 149-186, 1968.

[24]P.K.Feyerabend, *Against Method*, Verso, London, 1975.

[25]F. Bonsignorio, J.Hallam and Angel P. Del Pobil (eds), GEM Guidelines, Euron GEM Sig Report, (2008)

[26]C. Ruhla, The Physics of Chance, Oxford University Press, Oxford, 1992.

[27]G. Boniolo, A Contextualized Approach to Biological Explanation, Philosophy, 80, 219-247, 2005.

[28]E. Nagel, The Structure of Science. Problems in the logic of the scientific explanation, Harcourt, New York, 1961.

[29]S.D.Mitchell, Dimensions of Scientific law, Philosophy of Science, 67, 242-265, 1997.

[30]N.Goodman, Fact, Fiction and Forecast, Harvard University Press , Cambridge (MA, USA), 1954.

[31]C.G. Hempel, Aspects of Scientific Explanation and other essays in the Philosophy of Science, Free Press, New York, 1965.

[32]C.G. Hempel, P. Oppenheim, Studies in the logic of explanation, Philosophy of science, 15, 1948.

[33]M. Schlick, Die causalitaet in den gegerwaertigen physik, Die naturwissenschaften, 19, 145-162, 1931.