Robotic Competitions as Experiments: A Critical View

Francesco Amigoni Andrea Bonarini Giulio Fontana Matteo Matteucci Viola Schiaffonati

EXTENDED ABSTRACT

In recent years, a point of view that considers robotic competitions as experiments has emerged [1, 2, 3, 4], and represents the core of at least two EU projects: RoCKIn[1] and euRathlon[2]. This trend originates in the attempt of making methodologies in autonomous robotics sounder and more rigorous while, at the same time, exploiting research and infrastructures for competitions that the robotics community has developed during the years. Obviously, between competitions and experiments there are several differences: the most notable ones are probably that an experiment evaluates a specific hypothesis while a competition usually evaluates general abilities of robotic systems and that competitions push to development of solutions, while experiments aim at exploring phenomena and sharing results. Nonetheless, there are a number of reasons for recasting robotics competitions as experiments, considering traditional experimental principles (comparison, repeatability, reproducibility, generalization, ...) as guidelines.

Competitions usually involve some robots in a dynamic, but rather controlled, environment and, having clear measures of success, provide opportunities to benchmark approaches against each other. Furthermore, they require integrated implementation of complete robotic systems, promoting a new experimental paradigm trying to integrate the rigorous evaluation of specific modules in isolation (typical of robotics research). This experiment-oriented perspective on competitions not only can help better merge research with demonstrations, but can also provide a common ground for comparison of different solutions. Reframing competitions as experiments increases their scientific rigour while trying to maintain their distinctive aspects: competitions are appealing (people like to compete) and they take place with regularity and precise timing, showcasing the current state-of-the-art in research/industry. Finally, competitions promote critical analysis of experiments out of labs and they share among participants the cost and effort of setting up complex experimental installations.

However, the relationships between competitions and experiments are multifaced and, in this work, we attempt at taking a more articulated view on this issue. In particular, we aim at investigating the conditions under which competitions could be considered as a particular form of experimentation. Although competitions can be considered as a way of comparing the performance of robots, their character of one-time demonstrations puts some limits on the generalizability and replicability of their results and do not necessarily prove that some robotic systems are better than others. As it has been already noticed [5], robotic competitions are not necessarily experimental procedures, unless the robots are specifically designed to test the effectiveness of particular aspects of robots clearly expressed by an experimental hypothesis.

In computer science and engineering the concept of experiment is used in a variety of ways, and the label *experimental computer science* has been differently interpreted in the development of the discipline. Generally, experiments in computing can be intended as the empirical practice to gain and check knowledge about a system. However, this idea can be conceptualized in different ways. Among the recent conceptualizations we believe that the one proposed by [6] covers in an appropriate way the current practice by proposing five different meanings for an experiment, listed in rough order to increasing complexity and sophistication.

- *Feasibility experiment.* It is the loosest use of the term experiment that can be found in many works reporting and describing new techniques and tools. Typically, the term experiment is used in this case with the meaning of empirical demonstration, intended as an existence of proof of the ability to build a tool or a system.
- *Trial experiment.* This is a step further than the feasibility experiment, as it requires the evaluation of various aspects of a system using some predefined variables which are often measured in laboratories, but can occur also in real contexts of use (given some limitations).
- *Field experiment.* It is similar to trial experiment in its aim of evaluating the performances of a system against some measures, but it takes place outside the laboratory in complex sociotechnical contexts of use. The system under investigation is thus tested in a live environment and features such as performance, usability, or robustness, are measured.
- *Comparison experiment.* In this case the term experiment refers to comparing different solutions with the goal of looking for the best solution of a specific problem. Typically, comparison is made in some setup and is based on some measures and criteria to assess the performance. Thus alternative systems are compared and, to make this comparison as rigorous as possible, standard tests and publicly available data have been introduced.

All authors are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Italy francesco.amigoni@polimi.it

[1] http://rockinrobotchallenge.eu

[2] http://www.eurathlon.eu/site/

- *Controlled experiment.* It is the golden standard of experimentation of traditional scientific disciplines and refers to the original idea of experiment as controlled experience, where the activity of rigorously controlling (by implementing experimental principles such as reproducibility or repeatability) the factors that are under investigation is central, while eliminating the confounding factors, and allowing for generalization and prediction.

Note that, even if these types of experiments usually are mixed up in daily research activities performed in computing, they give us the idea of the range of the different experimental practices. In the following, we assume these types of experiments as significant for robotics, too.

We now discuss whether and how some robotic competitions can be mapped to the above types of experiments and we consider, in particular, what the features are, if any, that make these competitions experiments of that kind. Let us start from RoboCup and consider, for example, the competitions in the Middle Size Soccer League[3], in which two robotic teams play against each other in a soccer game. These competitions can be clearly considered as particular feasibility experiments and, partly, as trial experiments given that they can evaluate various aspects of a system using predefined variables. However, notwithstanding the fact that they involve two competing robot teams, it is harder to consider the competitions as instances of comparison experiments. While the settings and the parameters that define the competitions are usually very well specified, the measures and the criteria according to which the two robotic systems (teams) are compared are clearly defined only for the purposes of the game and not expressed in a general way by means, for example, of an experimental hypothesis. It is therefore difficult to generalize any conclusion about the general behavior of robots and their components from the fact that one team won, say, 2-0 against a second team.

Let us continue with RoCKIn, an EU project funded under FP7, that intends to provide a more principled approach to define and evaluate competition results. In both @Home and @Work competitions, it is clear the attempt to move towards comparison and controlled experiments. For example, one of the main features of the RoCKIn competition is the presence of two classes of benchmarks, called *task benchmarks* and *functionality benchmarks*. The first ones are devoted to evaluating the performance of integrated robotic systems, while the second ones focus on the performance of specific sub-systems (like object recognition and localization). A task benchmark deals with complete robot systems, implying that a large set of interacting robot elements are examined together at the same time. Functionality benchmarks, on the contrary, try to shed some light on the dependencies between sub-systems and the whole robotic system defining a precise setup in which a single robot functionality can be evaluated. Such evaluation is performed according to well specified quantitative measures and criteria, specific for the functionality under test.

[3] `http://wiki.robocup.org/wiki/Middle_Size_League`

These two examples show us already how the relationships between competitions and experiments greatly vary from case to case. A more complete theoretical work and a more extensive practical experience will be necessary to determine whether, and under what conditions, real-world robot competitions can be considered as scientific experiments. It seems that the notion of control should play a decisive role, as controlling the experimental factors that are to be investigated constitutes one of the key factors of the experimental method. Moreover, in particular when considering the engineering sciences, the presence of an epistemological aim seems crucial to call something properly an experiment: experiments are done to *learn* something, where this learning does not need to be the primary objective of doing an experiment, but can be also a secondary goal. It seems that transporting the many attractive aspects of competitions in more scientific contexts is a promising approach, which deserves to be further investigated with a critical approach that does not give for granted the many similarities between competitions and experiments, but reflects on the conditions under which equating competitions and experiments is reasonable and profitable.

REFERENCES

[1] M. Anderson, O. Jenkins, and S. Osentoski. Recasting robotics challenges as experiments. *IEEE Robotics Automation Magazine*, 18(2):10–11, 2011.

[2] R. Cohn, A.G.and Dechter and G. Lakemeyer. The competition section: a new paper category. *Artificial Intelligence*, 175:iii, 2011.

[3] D. Holz, L. Iocchi, and T. van der Zant. Benchmarking intelligent service robots through scientific competitions: The RoboCup@Home approach. In *Proc. AAAI Spring Symposium on Designing Intelligent Robots: Reintegrating AI II*, pages 27–32, 2013.

[4] B. Smart. Competitions, challenges, or journal papers. *IEEE Robotics Automation Magazine*, 19(1):14, 2012.

[5] L. Takayama. Toward a science of robotics: Goals and standards for experimental research. In *Robotics: Science and Systems (RSS) Workshop on Good Experimental Methodology in Robotic*, Seattle, WA, 2009.

[6] M. Tedre and N. Moisseinen. Experiments in computing: A survey. *The Scientific World Journal*, Volume 2014:1–11, 2014.