

Robotics Competitions as Experiments: a Critical Review



ROCKIn



F. Amigoni, A. Bonarini, G. Fontana, M. Matteucci, V. Schiaffonati (Politecnico di Milano)
Artificial Intelligence and Robotics (AIRO) 2014, Pisa

Presentation Purposes

1. Introduce a critical perspective on considering robotics competition as scientific experiments inspired by a taxonomy of experiments in computer science
2. Present the FP7 funded RoCKIn competitions to describe how their design has been inspired by the aim for benchmarking and experimental validation

Competitions lead to...

- Innovation
 - powerful means to foster progress in R&D
 - introduce/compare best practices
 - lead to technology transfers
- Focused R&D
 - research issues derived from real-world problems
 - development of commonly accepted test beds and benchmarks
 - experimental validation of state-of-the-art research
- Awareness of new technologies among citizens
 - increase the attractiveness of scientific & engineering disciplines
 - provides hand-on opportunities for primary/high school education (and university level education too)



EU funded Challenges/Competitions

- “RoCKIn is an EU project that will be run over next three years, consisting of robot competitions, [...] facilitating cognitive and networked robot systems’ testing, and streamlining research and development through standardized testbeds and benchmarks”



- “euRathlon is a new outdoor robotics competition which invites teams to test the intelligence and autonomy of their robots in realistic mock emergency-response scenarios”



- ... ADD HERE EUROC



Competition as Experiments

- In recent years, a point of view considering competitions as experiments has emerged

“Challenge and competition events in robotics provide an excellent vehicle for advancing the state of the art and evaluating new algorithms and techniques in the context of a common problem domain.

[...] treat competitions and challenges as repeatable experiments.”

Monica Anderson, Odest Chadwicke Jenkins, and Sarah Osentoski
“*Recasting Robotics Challenges as Experiments*”

Competition as Experiments

- In recent years, a point of view considering competitions as experiments has emerged

“One-time demonstrations of robot performance (e.g., grand challenges or other competitions) in robotics are one way of comparing the performance of robots, but they do not necessarily prove that one’s robotics research is consistently better or worse than another lab’s.”

Leila Takayama (Google[x], formerly at Willow Garage)

“Towards a Science of Robotics: Goals and Standards for Experimental Research”

At a First Sight...

- Competitions (and Challenges) share some similarities with Experiments ... but also have some differences

Competitions	Experiments	
Precisely definite settings	Controlled conditions	✓
Measuring performance: scoring	Measures and criteria	✓
Evaluating a whole robot system	Evaluating just one (or few) robot ability	✗
Hardly repeatable in the same conditions (e.g., challenges)	Intended to be repeatable in the same conditions	✗

Leila Takayama (Google[x], formerly at Willow Garage)

“Towards a Science of Robotics: Goals and Standards for Experimental Research”



Purpose #1

- Lets take a deeper critical view on the relationships between competitions and experiments
- One of our main insights is that **different competitions map to different types of experiments**
- Agenda:
 - Definition of terms: competition and experiment
 - Examples of mapping from competitions to experiments
 - Conclusions

Competitions

- “The act or process of trying to get or win something (such as a prize or a higher level of success) that someone else is also trying to get or win”
(Merriam-Webster)
- A robot competition usually involves:
 - some robots
 - a dynamic, but rather controlled, environment
 - clear measures of success

Experiments

- An experiment is a controlled experience, namely a set of observations and actions, performed in a controlled context, to test a given hypothesis
- In robotics, experiments are the *rigorous* empirical practice to gain and check knowledge about a system
- Some principles are usually ascribed to experiments: replicability, reproducibility, comparison, generalization, ...
- How are experiments intended and employed in computing?

Tedre and Moissenenen,
“Experiments in computing: a survey”, The Scientific World Journal, 2014.

Types of Experiments

- **Feasibility experiment:** empirical demonstration, existence of proof of the ability to build a tool or a system
- **Trial experiment:** evaluation of various aspects of a system using some predefined variables which are often measured in laboratories
- **Field experiment:** evaluating the performances of a system against some measures, outside the laboratory in complex sociotechnical contexts
- **Comparison experiment:** comparing different solutions with the goal of looking for the best solution of a specific problem; comparison is made in some setup and is based on some measures and criteria to assess the performance
- **Controlled experiment:** the golden standard of experimentation of traditional scientific disciplines, refers to the original experiment idea as controlled experience, where the activity of rigorously controlling the factors that are under investigation is central, while eliminating confounding factors, and allowing for generalization and prediction

RoboCup MSL Soccer League

- Two robotic teams play against each other in a soccer game
- The environment precisely defined and can be easily reproduced (not true for the opponent team and some conditions, such as light and environmental noise)
- 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
- This competition can be considered as a feasibility experiment and, partly, as a trial experiment



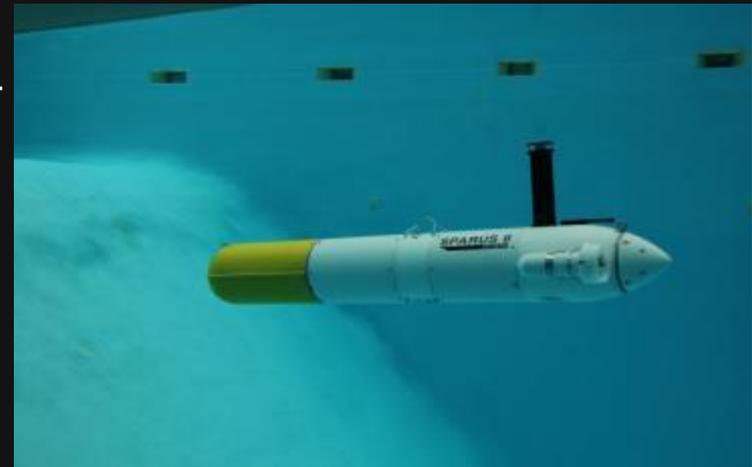
DARPA Robotics Challenge (DRC)

- The DRC consists of tasks related to human assistance in responding to disasters, “It was designed to be extremely difficult”
- Tasks settings are precisely defined (even if we could not find any official document about it)
- Task completion is evaluated using thresholds (e.g., number of open valves) plus bonus points. Time is not a factor (taking 30 seconds and 30 minutes to complete a task is worth the same amount of points), but is used as tiebreaker
- This competition could work as a **field experiment**



euRathlon

- euRathlon2013 has been about land robotics, while euRathlon2014 about underwater robotics
- Settings and tasks are defined precisely. “Competitors have the possibility to deal with real life conditions, e.g., limited visibility”
- Scores are a mix between measured quantities and subjective judgments given by a Judging Team
- This competition seems to move from **trial to field experiment**, not actually providing evidence that it can work also as a comparison experiment



Humanitarian Robotics and Automation Technology Challenge

- Robots explore the environment performing autonomous landmine detection
- Arena and the mines specifications are defined in the rules and can be easily reproduced. There are both simulation and real world runs
- Scoring combines the number of correctly detected (and unexploded!) mines, the area swept, ...
- This competition could be considered as standing between **trial and field experiments**



RoCKIn@Work

- RoCKIn@Home focuses on domestic service robots that have to perform socially useful tasks
- Home settings are precisely defined in the rules through
 - *Task benchmarks*: evaluating the performance of integrated robotic systems in performing tasks
 - *Functionality benchmarks*: evaluating the performance of specific sub-systems (like object recognition and localization)
- Evaluation is performed according to specified quantitative measures and criteria.
- This competition aims at being a **comparison experiment**



RoCKIn@Work

- RoCKIn@Work focuses on innovative robot applications in industry
- Factory settings are precisely defined in the rules through
 - *Task benchmarks*: evaluating the performance of integrated robotic systems in performing tasks
 - *Functionality benchmarks*: evaluating the performance of specific sub-systems (like object recognition and localization) and a, de facto standard platform is used
- Evaluation is performed according to specified quantitative measures and criteria
- This competition comes closer to comparison experiments



Summary (not a ranking!)

Humanitarian Robotics and
Automation Technology
Challenge

euRathlon

RoboCup Middle
Size Soccer League

DARPA
Robotics
Challenge

RoCKIn

Feasibility
experiment

Trial
experiment

Field
experiment

Comparison
experiment

Controlled
experiment

RoCKIn

Conclusions #1

- Motivating question: if, and under what conditions, can real-world robot competitions be considered as scientific experiments?
- No definitive answer has been provided, but some examples have been discussed which cover a wide range of experiment types
- A more complete theoretical work and a more extensive classification of existing competitions are needed
- Transporting the many attractive aspects of competitions in more scientific contexts is a promising approach, which deserves to be further investigated

Purpose #2

- Lets take a deeper look into the RoCKIn competitions design
- One of our main claim is that RoCKIn competitions have been designed as scientific competitions
- Agenda:
 - Brief Description of RoCKIn Comeptitions
 - Design of the RoCKIn@Home competition
 - Conclusions

Driving Robotics Forward through Collaboration, Benchmarking and Competitions

The logo for the ROCKIn project, featuring the word "ROCKIn" in a stylized, metallic, 3D font. The letters are dark grey with highlights and shadows, giving them a mechanical appearance. The background is a blue circuit board with various electronic components and labels like "TYT", "SF1", "A12", "30", "31", "38V", "C12", and "-220V".

ROCKIn



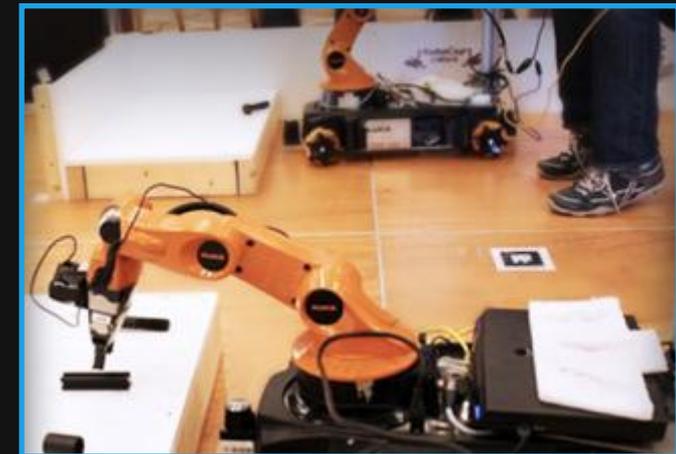
This project is funded by
the European Union



Matteo Matteucci (Politecnico di Milano)
Forum on Robotics Challenges & Competitions
- ICRA 2014, Hong Kong -

RoCKIn@Work

- **Innovative robot applications in industry that**
 - work interactively with humans
 - have reduced initial programming requirements
 - have enhanced physics simulation capabilities
- **Contribute to the continued commercial competitiveness of European industry**



RoCKIn

RoCKIn@Home

- **Socially beneficial domestic service robots that**
 - have enhanced networking and cognitive abilities
 - support the impaired and the elderly
- **Contribute to an improved quality of life for the population of Europe**



ROCKIN

Competitions as Experiments

- Benchmarking activities in RoCKIn are aimed at
 - Inspiring the design of competitions to allow for benchmarking
 - Designing suitable metrics for the competitions (and for robotics)
 - Applying such metrics during the competitions
 - Comparing results after the two competitions
- Experimental method suggests experiments to allow:
 - Comparison
 - Reproducibility / repeatability
 - Justification / explanation

Functional and task benchmarks

- Competitions may challenge robots at two different levels (ability vs capability in SRA?)
 - Task Level: evaluation of whole systems on a specific task (e.g., the “bring me the glasses” tasks)
 - Functionality Level: evaluation of modules implementing, in a general manner, functionalities required by the competition tasks (e.g., grasping and manipulation)
- Benchmarking competitions should allow independent evaluation at both levels
 - To encourage participation of people interested in specific aspects of robotics (e.g., object recognition)
 - To evaluate at what extent the interplay among modules is relevant (e.g., the precision in positioning before grasping)

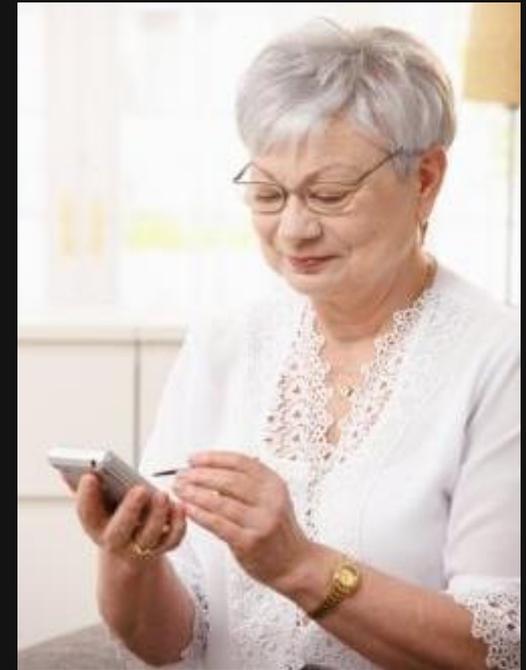
User Story: Granny Annie

Granny Annie is a nice but slightly seasoned lady. Luckily, she could get into a new program, sponsored by her health & social security insurances, by which elderly people are supplied with household and elderly care robots to assist in managing and mastering their daily lives.

- Task: “Getting to know my home”

Granny Annie is waking up and today she feels a bit tired because she has not slept very well. Still a number of tasks need to taken care of. The home robot will help her in all these tasks.

- Task: “Catering for Granny Annie’s comfort”
- Task: “Welcoming visitors”
- ...



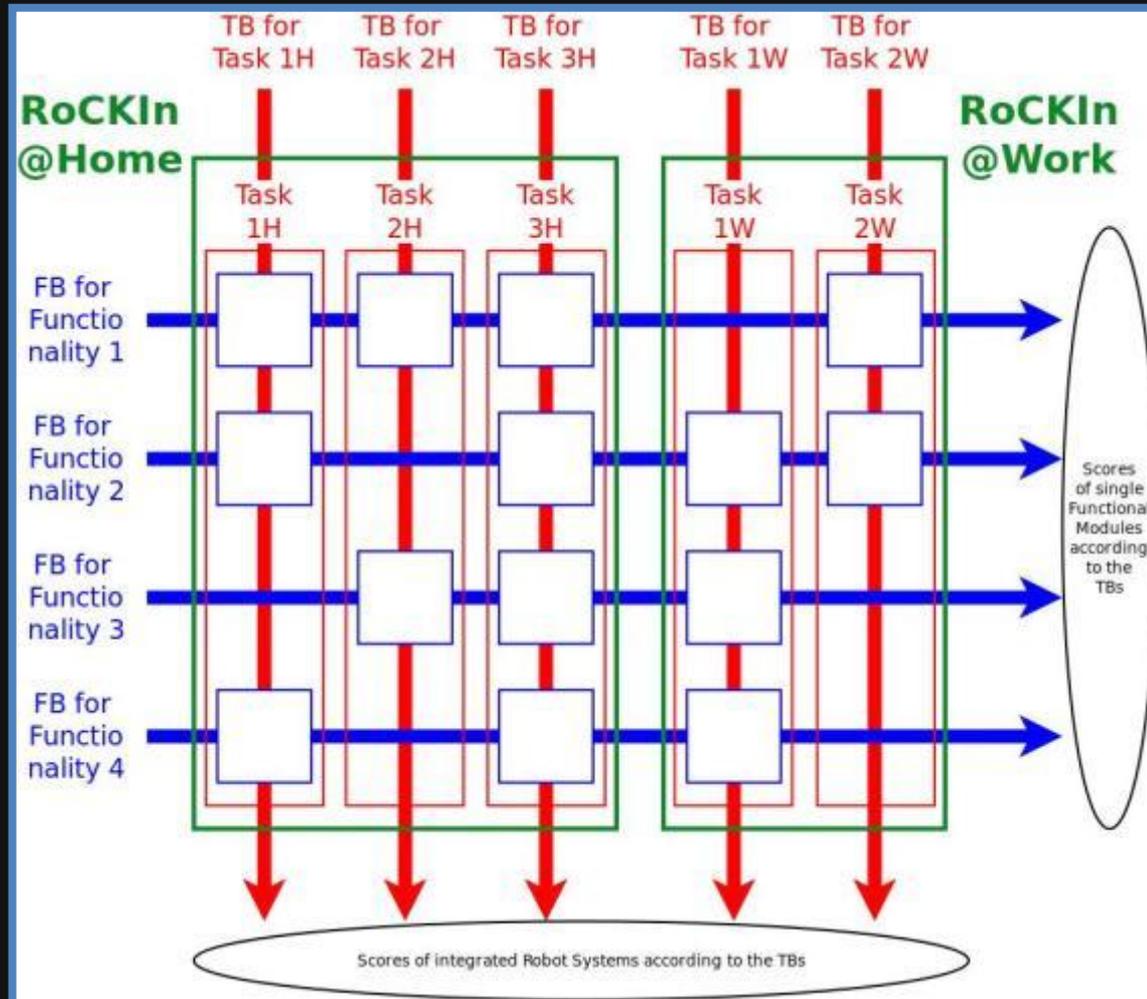
RoCKIn@Home Tasks & Functionalities

- Tasks Benchmarks
 - Comfort Providing Task Benchmark
 - Visitor Handling Task Benchmark
 - Environment Learning Task Benchmark
- Functionality Benchmarks
 - Object Perception Functionality Benchmark
 - Object Manipulation Functionality Benchmark
 - Speech Recognition Functionality Benchmark



ROCKIN

A big picture ...



RoCKIn Competition 2014

La Cité de L'Espace, Toulouse, on 26-30 November 2014
(during the 2014 European Robotics Week / ERW 2014)

- 10 teams participated (3 @Work, 7 @Home)
- Co-located events:
 - research workshop on HRI and
 - workshop industry-meets-academia



With the support from the City of Toulouse
and the Toulouse region

Co-organized by Rachid Alami/LAAS

ROCKIN

Teams and Benchmarks

Team	TBM-1	TBM-2	TBM-3	FBM - 1	FBM - 2	FBM - 3
b-it-bot@Home		X	X		X	X
BARC	X	X	X			X
Homer@UniKoblenz	X	X	X	X	X	X
Pumas@Home	X		X	X		X
SocRob@home		X	X			X
Ursus-Team				X		X
Watermelon Project		X		X	X	X

- Modules for benchmarking developed
 - Data from robot collected in Task Benchmarks
 - Data collected from Functional Benchmarks

Conclusion #2

- The classification of “Robotic Competition as Experiments” has been used to guide the design of the RoCKIn Competitions
- Data available so far do not allow a statistically significant analysis of performance, nevertheless they will allow a rigorous comparison with future RoCKIn Editions
- You are invited to participate at
 - the RoCKIn 2015 Field Exercise (Peccioli, March 18-22)
 - the RoCKIn 2015 Competition (Lisbon, fall 2015)