

**Refutable Robotics Research**  
**What we mean**

Fabio P. Bonsignorio  
Angel P. Del Pobil  
John Hallam

- 'Look Ma, No Hands' syndrome?
- Replication of experiments
- Performance measure benchmarks to allow results comparison
- Needed to foster research advancement and enable practical application of research achievements

As the complexity of developed robotic and intelligent systems grows, it is more and more needed to define proper experimental approaches and benchmarking procedures.

Trustable benchmarks are needed in order to allow the comparison of the many research results in service robotics research end enable their industrial application.

if robotics aims to be **serious** science replication of experiments deserves serious attention.

Are we really able 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?**(0)???**

New more successful implementations of concepts already presented in literature, but not implemented with exhaustive experimental methodology, risk to be ignored, if appropriate benchmarking procedures, allowing to compare the actual practical results with reference to standard accepted procedures, are not in

place. **Is this true?**

Both replication and benchmarking are needed to foster a cumulative advancement of our knowledge of intelligent physical agents and even to correctly appreciate disruptive innovation in the science ((1) ???) and technology of robots.

Should we take inspiration from biology and medicine (2) ???

If robotics aims to be serious science, serious attention must be paid to experimental method.

What is an 'experiment' in robotics?

Whether you see robotics as the science of intelligent physical agents ('embodied cognition') or as the branch of engineering that, through mechatronic integration, aims to build autonomous or semi-autonomous machines for many diverse tasks, it must be seen as a scientific quantitative discipline. What does this mean?

## *Replication&Falsification*

As it is known K.Popper defined in a very tight way the requisites for a discipline to be considered 'scientific'. In social science, management and economics exact repetition is often seen as a limit case, experiments that systematically vary one or more input parameters of a system under study to see whether its output parameters remain stable or change according to the expected model in a predictable way.



## *Replication&Falsification*

Only when the model fails clearly in a number of varied experimental setup it is considered 'not replicable'. Nevertheless, as already noticed, all disciplines aiming to be considered 'scientific' incorporate a concept of experiment replication and a concept of 'falsification' of theory through experiments.

## *Replication&Falsification*

There are different modulation of this concept, but whether we think we are in a cumulative phase in the development of a scientific field or in presence of a 'disruptive' creative paradigm shift, as somebody is claiming in nowadays robotics, a kind of widely accepted experimental methodology is needed in order to be able to ground the advancement of research on a shared quantitative language.

## *Replication & Falsification*

A **clinical trial protocol** is the detailed written plan of a clinical experiment.

It may be inspiring looking at the US NCI guidelines for drafting a clinical trial protocol: the emphasis on signaling 'adverse events', the definition of 'criteria for response assessment', the necessity of defining clearly principal and secondary hypotheses to be validated.

The **statistical section** of the protocol is asked to define how the data will be analyzed in relation to each of the objectives.

In particular it expects that an acceptable trial specify, with reference to the study objectives:

- Method of randomization and stratification
- Total sample size justified for adequate testing of primary and secondary hypotheses
- Error levels (alpha and beta)
- Differences to be detected for comparative studies
- Size of the confidence interval of the estimates.

## ***Replication & Falsification***

It seems clear that in robotics the experimental methodology standards are currently in many cases weaker, and the syndrome 'it worked once, in my lab' could be more widespread than we may think.

## *Replication & Falsification*

A limit to replication is given by the huge variability of robot machines.

Perhaps, following the biomedical analogy, we have to compare behaviors and performances of different 'animals'.

## *Discussion*

Why we need both replication AND benchmarking?  
FACT CHECK: Benchmarking is more studied than  
Replication

- SLAM
- Mobile Robots' Motion Control
- Robot Obstacle Avoidance
- Grasping
- Visual Servoing
- Autonomy/Cognitive tasks

***Discussion***

**Good Old Fashioned Robotics Research?**

**New AI and Robotics**

**Mechanics and Robotics**

**Control theory and Robotics**

**Cognition Sciences and Robotics**



## *Discussion*

### Replication of robotics experiments

- Research Reporting in Biology and Medicine
- Evidence Based Medicine
- Early studies in robotics research replication

***Discussion***

**Refutable Robotics Research**

**What we mean**

**Open issues**

**The road ahead**

## *Discussion*

The bare replication of experiments and the quantitative comparison of research results in robotics raise many challenging issues.

This is due to the variety of applications, tasks, mechanical structures, sensor sets, actuators, control system, software architectures, required levels of flexibility and autonomy, and so on.

### *Discussion*

In some 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 [Chatila,2006], entropy measures are used to rank environment complexity, with reference to the navigation task.

## Does this help?

### *Call to Arms*

There is a widespread perception of the need of improving experimental practices in robotics, among many others world wide initiatives, the Euron SIG GEM(II and 3) is trying to address these needs.

It is thought that proper and widely accepted replication procedures and performance benchmarks are needed to allow the cumulative progress of robotic science and technologies and even to assess the value of new disruptive ideas.

(RSS09 Workshop on GEM, too!)

**Thank you!**

Fabio P. Bonsignorio  
fabio.bonsignorio@heronrobots.com  
Angel P. Del Pobil  
pobil@icc.uji.es  
John Hallam  
john@mmmi.sdu.dk