

Robotic research: are we applying the scientific method?

Gianluca Antonelli*

Department of Electrical and Information Engineering, University of Cassino and Southern Lazio, Cassino, Italy

Keywords: robotic control systems, benchmarking in robotics, scientific method, research validation, robotic research

Once upon a time validation of robotic research was relatively straightforward. Let us assume, for example, that a researcher had published in a journal a novel adaptive control law with a numerical example on a two-link robot. Beyond the formal proof of convergence, he supplied to the reader the differential equations used to model the system, including the corresponding dynamic parameters (no more than 20 numbers), the eventual quantization and discretization of the controller, the solver details of the software used, and the sensor noise statistics. Not only the reviewers, thus, but also each single reader would have the possibility to re-run the numerical simulations in a half-day of work. The community would have the possibility to test, validate, generalize, and benchmark the algorithm.

Since then, robotics has changed, the machines are now much more complex in their kinematics, number of degrees of freedom, and are filled with several sensors. Also, giant steps have been made: the robots left the confined industrial cells to jump within unstructured environments, not only in the industry but also in the houses, the museums, the airports, and the post-disaster sites; they perform a number of exciting tasks such as exploration, maintenance, interaction with humans, search and rescue . . . wait, is it really so? Beyond specific outstanding experiences, beyond the claims of the constructors and lab's directors, how many robots run, autonomously or semi-autonomously, in our daily lives? Not so many, to be honest. A few vacuum cleaning robots, this is all (Guizzo, 2015). While we have several noticeable robotic tools (parking assist systems, lane keeping assist systems, space systems), where are all the learning and adaptable robot protagonists of thousands of scientific publications in the last years? Our robots can avoid the *predicted unpredicted* events, but what about the *unpredicted*? The information required to validate the two-link example above is obviously not possible any more but why are we experiencing so large a gap between claimed and real robotics? Why has it been the case for several years now that the robotics *revolution* is regularly postponed to the next 10 years . . . ?

The grand challenge for the robotics community is to discuss, from its foundations up, the way its research is conducted. It is a huge effort involving complex interactions among the institutions, the ministries, the funding agencies, and the individual researchers' careers. Research is funded by selection of proposals, at each call more and more imaginative which, however, most of the time end with more or less disappointing demos. This process includes perforce to review the validation of the research process in a wide sense and, within this, the publishing process. The latter is becoming (apparently) faster and more selective with new ideas spread out and absorbed by other researchers in a very short time during which a paper placed in the hands of a reviewer or a debatable reject may be a dramatic event.

The previous claims are intentionally provocative, and so is the title of this article: are we (still) applying the scientific method in robotics? Let us frankly discuss this question.

The Oxford English Dictionary defines the scientific method as "a method or procedure that has characterized natural science since the 17th century, consisting in systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses." The strict link existing with theory and experimentation is evident and further elaborated by, for instance, Karl Popper, who claimed that the criterion of the scientific status of a theory is its falsifiability,

OPEN ACCESS

Edited by:

Torsten Kroeger,
Google Inc., USA

Reviewed by:

Vincent Creuze,
University of Montpellier, France
Sylvain Calinon,
Idiap Research Institute, Switzerland

*Correspondence:

Gianluca Antonelli
antonelli@unicas.it

Specialty section:

This article was submitted to Robotic Control Systems, a section of the journal Frontiers in Robotics and AI

Received: 20 April 2015

Accepted: 19 May 2015

Published: 01 June 2015

Citation:

Antonelli G (2015) Robotic research: are we applying the scientific method? *Front. Robot. AI* 2:13. doi: 10.3389/frobt.2015.00013

or refutability, or testability (Popper, 1963). Within the same framework, we find the sentence attributed to Einstein – “No amount of experimentation can ever prove me right; a single experiment can prove me wrong” – and the famous Wolfgang Pauli saying, of an argument that fails to be scientific because it cannot be falsified by experiment, “it is not only not right, it is not even wrong!” The term “falsifiable” does not mean something is made false, but rather that there exists an observation or experiment that might prove it to be false. Yes, all our scientific theories are perpetually subject to falsification. What is not scientific falls within the categories of non-science or pseudoscience, with slightly different meanings and a certain negative attitude toward the term pseudoscience. Social sciences are often considered as non-science because of the impossibility to exactly replicate the experiments. Economy and medicine, for example, are characterized by the uniqueness of each single event and thus the impossibility to, for example, implement a successful macroeconomic approach from one country to another.

The scientific method is not universally recognized as the sole medium for advancing knowledge, Paul Feyerabend’s opinion is that epistemological anarchism is probably the only paradigm that allowed science and knowledge to reach their current levels (Feyerabend, 1993). He criticized any standardized methodology or any possible authority for science. This opinion stands on various considerations, one of which is that a large number of all scientific discoveries are estimated to have been stumbled upon, rather than sought out. Several prestigious scientists consider themselves *lucky*; Louis Pasteur is credited with the saying “Luck favors the prepared mind” and, more recently, Nassim Nicholas Taleb, a former market analyst, developed the concept of “anti-fragility” as the capability *to be ready to profit from chance*.

Within a robotic framework, the considerations above bring us toward two aspects: interaction and replication. We need, as a research community, to exchange our experiences, our algorithms, the pieces of code since we need to replicate the experiments, to find bugs or theoretical flaws, and to benchmark. Let us thus have a closer look at the robotics case.

The Robotic Case

Conferences coffee-breaks are well-known critical sources of information. Among the various chats, few sentences collected are as pertinent here as “I asked a Ph.D. student to implement the method of xyz but it never worked, we wasted 6 months and we do not know why,” or “We implemented the method of xyz but it appears to work only on full moon nights. We could not prove it and we never published the results.” It is clear that the community is wasting resources; we know that only formal *Comments on . . .* can be published for major analytical flows but it is quite hard to find papers commenting on the implementation issues (robustness, sensitivity to parameters, etc.). On the other hand, each of us experienced the issue of having asked a junior to implement an algorithm and, at its failure, to remain with the unpleasant flavor of not having additional resources to investigate why and not trusting both the junior and the algorithm.

A different sentence came from a senior colleague involved in robotic applications of social sciences that, pretending not to need

to follow the scientific method paradigm, finally claimed, “asking for formal proof is unfair, the world is not mathematics.” The former words, taken alone, might also be debated but, in a context of an escape-from-validation strategy, appear to be debatable. Sometimes, researchers from the social sciences raise as a flag a personal interpretation of the Gödel’s incompleteness theorems such that mathematics is once and for all kept out from discussion on their view of artificial life systems.

Another risk pushing the research away from the scientific method is the so called *a posteriori reasoning*, affecting a surprisingly large number of domains. Let us consider the classical example of the group of monkeys playing the stock market. Each morning, the monkeys forecast whether the market will rise or fall at the end of the day. Assuming a pure casual choice, it is reasonable to assume that they will split in half and thus half will be correct. The following day the same operation is repeated, still half of the monkeys will be correct; one-quarter of the initial set, moreover, will guess for two consecutive days. With a large enough set of monkeys, a certain number of them will guess for several days, or weeks, and will be considered as smart enough to interpret the stock market. The flaw is that only *a posteriori* we know which monkey will win. Similarly, the human body is considered as a perfect machine because it adapts perfectly to environmental conditions. This is one possibility but we cannot discern it from another possibility: the physical characteristics are spread totally casually among the species. The species which do not adapt to survive on this planet disappear and only few of them, among which are listed humans, happily survive. *A posteriori* a perfect design. . . . This risk strongly affects the area of behavioral robotics. Let us assume that we implement via software some kind of behavior that pushes the robot toward a goal while pulling it far from obstacles. We run the experiment and then, *a posteriori*, whatever happens is interpreted in the framework of our solid behavioral structure (“the robot is *attracted* by the food,” “it is *afraid* of the obstacle,” “it is *curious* . . .”). Let us increment the number of robots and behaviors; isn’t it similar to what is considered “emergence behavior”? Arkin (1998) wrote “Emergence is often invoked in an almost mystical sense regarding the capabilities of behavior-based systems. Emergent behavior implies a holistic capability where the sum is considerably greater than its parts. . . . The notion of emergence as a mystical phenomenon needs to be dispelled, but the concept in a well-defined sense can still be useful. . . . Coordination functions as defined in this chapter are algorithms and hence contain no surprise and possess no magical perspective.”

Words are important . . .

Etymology and Epistemology

Etymology is the origin and meaning of words, epistemology studies the nature and scope of knowledge. One interesting issue concerning etymology and epistemology has been raised by Heisenberg (1958), when he addressed the problem of describing new concepts with existing words. Heisenberg was walking at the edge of human knowledge and comprehension and thus his communicability trouble was deep and severe.

Surprisingly, in robotics, it seems that a similar problem is traversing the community. The following is a list of terms used in

recent scientific papers dealing with robots: intelligence, cognition, collective cognition, meta-cognition, smartness, consciousness, pre-consciousness, awareness, self-awareness, collective awareness, collective identity, collective memory, mood, emotion, and so forth (basically a two-combination of a huge set . . .). When reading those words we all are influenced by the etymology, of course, even if they are supposed to represent *new* concepts, basically algorithms. Given their algorithmic nature one would expect a clear definition of these terms, both in terms of a linguistic description and in terms of any analytical language (mathematics or code-like reasoning). However, a large part of the community is using these terms by circular definitions, i.e., by defining one using a second that uses a third and so on while the final term is defined using the first again, and intentionally refusing to deal with any formal description. Heisenberg claimed that a formal language, alone, would not have the same powerful communication capability of the human language; he was talking about the Copenhagen theory, however, and not of algorithms to be implemented in C.

The proportional-integral-derivative (PID) controller dates 1922 (Minorsky, 1922), do we need pages of dialectic to come out with always the same three actions? What is the difference between a master/slave architecture with respect to a father/son one? What is the added value for the yet-another-controller based on the optimization of a yet-another-functional in which the interpretation is given a new, fancy, name?

Interesting enough, the misuse of words is not specific to Robotics, as an example in Economy, there is a debate about the possible malicious use of the term *efficiency* (Buchanan, 2013).

This poses a big ethical issue since most of the foundations of the scientific method are neglected. On the other hand, this naturally conducts us to next Section dealing with the role of the media in robotics research.

The Role of the Media

There is a huge gap between the expectations of non-experts and what a robotic system can actually do. Often, scientific journalists push for impressive titles and videos while totally neglecting the scientific and technological aspects. The language used by the journalists is different from the one used by the research community and mainly aimed at achieving a larger impact rather than an adherence to reality. On the other hand, the funding agencies ask for a diffusion of the results on the general public in order to give evidence of the use of the public money. Multimedia interpretation of the results can be very informative, however, the temptation to use science-fiction terminology and to enrich the experimental video with post-processing editing is difficult to avoid, as it is clear by looking at the numerous attractive robotic cases and advertising-like projects' spots.

Replicability

The main road to follow the scientific method is to allow the replicability of the experiments. Two aspects need to be considered separately, perception and control. For the perception case, the real-time issue is not a critical variable to replicate or benchmark a certain algorithm. Raw data exist that allows the community to compare their own works. One of such examples is the RawSeeds

project funded by the European Community under the FP6 program¹ that permits the downloading of datasets composed by several sensors such as ultrasounds, inertial measurement units, vision and laser range finders, together with a given ground truth. Unfortunately, when dealing with control, the possibility to *exactly* replicate a certain experiment is much more difficult. To the best of our knowledge, in cases of control software being released to the community, and for obvious reasons, it is very rare that the hardware is provided together with the code. Even in that unlikely case, to replicate the experiment would involve a cost in terms of human resources that would discourage it. In the recent years, the adoption of common hardware and software platforms made it possible to reduce the differences among the laboratories and thus will probably have a positive impact on replicability.

In case it is necessary to replicate a robotics test, both for perception or control, involving humans makes the issue still more delicate. Similarly, to the other social sciences, human behavior is never the same in two consecutive tests, thus forbidding the possibility of an exact test replication. The next natural step is to move toward massive experimental campaigns and statistical analysis, but this brings into the play still additional problems, for example, considering the increased costs involved in large-dimensional surveys.

Competitions

One possible approach is to evaluate from the results, no matter what path and terminology has been used to achieve them, within the same environmental situation, i.e., in the same moment with the same constraints. We are talking about robotic competitions. Currently, there are dozen of them worldwide with different structures, such as the DARPA challenge in USA² or EuRoC³, Eurathlon⁴, and RockIn⁵ in Europe or the Mohamed Bin Zayed International Robotics Challenge in Abu Dhabi, plus several others involving mainly college student teams.

The competitions differ from each other in several aspects, including the possible money prize or the eventual financial contribution to the participating teams. In all the competitions, there is always a trade-off between objective metrics and a peer review aspect. In EuRoC, for example, the Consortium is in charge also of hosting the hardware platforms and providing the teams with more or less basic robotic functionalities. In a way, the teams are leveled and the focus is on the algorithms and applications.

The critical aspects concerning competitions are the motivation that pushes the researchers in participating. When the money prize is the sole appeal, the competition will probably attract teams with a commitment proportional to the amount of money. However, if competitions are the tool to benchmark and evaluate and to allow interaction within the community other captivating strategies should be envisaged. Also, competitions embeds an inherent risk of wasting effort in quick-and-dirty fixes employed to match the realms and deadlines of the challenges, which could potentially have an impact on longer term research.

¹<http://www.rawseeds.org>

²<http://www.darpa.mil>

³<http://www.euroc-project.eu>

⁴<http://www.eurathlon.eu>

⁵<http://rockinrobotchallenge.eu>

The Grand Challenge

Will the future bring an independent institution receiving funding to falsify the theories or to provide a standard for robotic algorithms? It is clear that research on research evaluation is a topic itself; it is dynamic and it changes with time. Even research in robotics, thus, is entering a new, challenging phase, for which we might need to refine our assessed evaluation convictions and even our beliefs and try to adapt our work to new environmental

conditions. It is probably the moment to enlarge the amount of people that can provide feedback on scientific publications by providing access to all and by including a metrics of the impact of the papers *after* their publication, following a technical-only threshold within a solid validation discussion. Open access, open review, competitions, full disclosure policies can be viewed in the perspective of trying to level the playing field amongst *competitors* in the research adventure, with the ultimate conviction being that this will definitively provide new and exciting results.

References

- Arkin, R. (1998). *Behavior-Based Robotics*. Cambridge, MA: The MIT Press.
- Buchanan, M. (2013). *Forecast: What Physics, Meteorology, and the Natural Sciences can Teach Us About Economics*. Bloomsbury Publishing
- Feyerabend, P. (1993). *Against Method*. Verso.
- Guizzo, E. (2015). "A robot in the family," in *IEEE Spectrum*, 52, 28–58. doi:10.1109/MSPEC.2015.6995630
- Heisenberg, W. (1958). *Physics and Philosophy*. New York, NY: Harper and Row.
- Minorsky, N. (1922). Directional stability of automatically steered bodies. *J. Am. Soc. Nav. Eng.* 34, 280–309. doi:10.1111/j.1559-3584.1922.tb04958.x

Popper, K. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Routledge.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Antonelli. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.