

On Research Reproducibility: An Interview with Gianluca Setti

By Eugenio Guglielmelli

Gianluca Setti (GS) is a professor in the Department of Engineering at the University of Ferrara, Italy, where he teaches circuit theory, analog electronics, and statistical signal processing. He has held several positions as a visiting professor/scientist, such as at École Polytechnique Fédérale de Lausanne, Switzerland (2002, 2005); the University of California, San Diego (2004); IBM T.J. Watson Laboratories (2004, 2007); and the University of Washington, Seattle (2008, 2010). He is also a permanent faculty member of Advanced Research Center on Electronic Systems (ARCES), University of Bologna. His research interests include nonlinear circuits, implementation and application of chaotic circuits and systems, statistical signal processing and compressive sensing, electromagnetic compatibility, and biomedical circuits and systems.

Dr. Setti received the 2013 IEEE Circuits and Systems Society (CASS) Meritorious Service Award, the 2004 IEEE CASS Darlington Award, and the 2013 IEEE CASS Guillemin-Cauer Award as well as the best paper award at the 2005 European Conference on Circuit Theory and Design (ECCTD 2005) and the best student paper award at the 16th International Zurich Symposium and Technical Exhibition on Electromagnetic Compatibility (EMCZurich 2005) and at IEEE International Symposium on Circuits and Systems (ISCAS 2011).

He has also held several editorial positions for the IEEE, including editor-in-

chief of *IEEE Transactions on Circuits and Systems—Part II* (2006–2007) and *IEEE Transactions on Circuits and Systems—Part I* (2008–2009); he served as a member of the editorial board of *IEEE Access* (2013–2015) and *Proceedings of the IEEE* (since 2015). He also served as the 2010 CASS president, and, in 2013–2014, he was the first IEEE Vice President for Publication Services and Products from outside North America. Dr. Setti was the technical program cochair of Nonlinear Dynamics of Electronic Systems (NDES 2000) (Catania, Italy), ISCAS 2007 (New Orleans, Louisiana), ISCAS 2008 (Seattle, Washington), IEEE International Conference on Electronics, Circuits and Systems (ICECS 2012) (Seville, Spain), and Biomedical Circuits and Systems Conference (BioCAS 2013) (Rotterdam, The Netherlands) as well as the general cochair of the International Symposium on Nonlinear Theory and its Applications (NOLTA 2006) (Bologna, Italy).



Gianluca Setti.

Q: What is your reference definition of research reproducibility (RR)?

GS: There are several definitions of RR since the concept is usually specialized and varies for each particular research community. In rough terms, and using the definition of reproducibility reported by *Wikipedia* [1], RR is linked to the idea that, as the ultimate product of academic research, papers

must be accompanied by the details of the full computational/experimental environment used to produce the results in the manuscript. Knowledge of this environment can be used to reproduce the results achieved by others and will help obtaining new discoveries based on those. I personally also like the definition by Buckheit and Donoho [2], “An article [...] in a scientific publication is not the scholarship itself; it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of

instructions” that generated the results. In other words, RR involves the complete knowledge of the data, the algorithms, the code, and the detailed experimental methods that were used to obtain all the results presented in an article. Another way to see this is that data, algorithms, and code are not simply ancillary information, but first class scholarly products as important as the paper itself.

Q: Are there different kinds of reproducibility?

GS: There are certainly different kinds of reproducibility depending on the particular area of science one deals with. Furthermore, the capability to guarantee reproducibility of research is more or less difficult depending on the

(continued on page 190)

Turning Point (continued from p. 192)

specific scientific subject. For instance, for mathematical sciences, ensuring reproducibility is quite straightforward since it requires clear reasoning and a

The concept is usually specialized and varies for each particular research community.

complete proof of the presented results, which can be (easily) followed by the readers. Once this is done, the results become

part of the common body of knowledge for the community and do not need to be proven again.

Conversely, for empirical sciences (such as life sciences, social sciences, and, in the area of interest of the IEEE, e.g., devices/circuits/systems implementation and characterization), the quest for RR involves, at a minimum,

- 1) a clear description of the methodology followed in a particular study or experiment
- 2) a detailed explanation of the laboratory procedures/protocols used
- 3) a thorough statistical analysis of the results obtained, highlighting their significance
- 4) the complete sharing of the data associated with the study/experiment
- 5) the sharing of the code and the features of the run-time environment that has (possibly) been used to produce the data.

The extent of this list clearly shows the intrinsic difficulties in guaranteeing reproducibility in this context.

For the computational sciences, which are a subject of interest for several scientific communities within the IEEE, guaranteeing RR mainly involves points 1), 4), and 5), thus resulting in an intermediate difficulty level.

Q: Why is RR important?

GS: Since the times of Galileo and Boyle, the basis of science has been the capability to replicate the results produced by other researchers, to build on their discoveries, to advance knowledge and technology. In other words, reproducing previous results to show the advantages of the proposed innovative

methodologies or techniques has always been the key to progress in science. Using Isaac Newton's famous expression, one can summarize this concept: "If I have seen further, it is by standing on the shoulders of giants" [3].

While this scientific approach worked remarkably well for centuries because of the ability of the scientific community to discover and correct mistakes and refine or completely change flawed theories and erroneous methodologies, in recent years something seems to have gone wrong in the self-correcting mechanism of science, particularly in the area of life sciences. Even if one does not consider the most outrageous cases of fraudulent research, such as the famous stem-cell scandal (which involved the retraction of two papers published in 2004 and 2005 in *Science* [4]), several recent studies have highlighted the impossibility of reproducing the results published in the vast majority of the papers under investigation. More precisely, and by way of example, according to Begley and Ellis [5], only 11% of 53 studies in the area of preclinical cancer drugs were reproducible, while Ioannidis et al. [6] show that this was also true for two of 18 papers in bioinformatics. What is worse is that similar findings have made their way into the general public press [7] and generated in the public opinion an increased sense of unease with respect to the way in which science operates.

A systematic adoption of RR practices is certainly necessary to reverse this worrisome trend. At the same time, its implications are far more important than this. RR is, in fact, fundamental since the following hold.

- It will foster growth in the capabilities for collaboration among scientists, which will help to overcome the increasing challenges posed by the rising number of multidisciplinary collaborations.
- It will produce an increase in the rate of innovation: researchers will advance technology more easily, and practitio-

ners will develop new products faster. This is, of course, a future that every scientist and practitioner will welcome as important steps forward for humanity.

Q: What is RR for the IEEE, and why it is important?

GS: IEEE is first and foremost a professional organization, and its publishing enterprise exists as a service to the community. One of the reasons reproducibility is important for the institute is that there are more indications that RR may actually soon be incentivized (if not mandated) by funding agencies in a similar way with respect to what has happened in recent years for open access. Another reason underlying its importance is that RR may simply become a more pressing request by the IEEE members and authors. More scientists are, in fact, interested in increasing the visibility of their discoveries: preliminary studies show a greater impact for those scientific works that share supplemental material together with the paper itself [8].

Consequently, simply because (part of) the IEEE community will need it, the development of an infrastructure supporting RR (at least in terms of storing/reusing data, code, and algorithms) may become, in my view, a pressing need for the IEEE in the not so distant future.

There are, however, other advantages that the adoption of RR will offer to IEEE. First, as previously mentioned, because of RR, the information made available through the IEEE conferences and publications will be more visible and directly usable by both scientists and practicing engineers. Furthermore, the adoption of RR will help the readers to navigate the large quantity of papers available on a specific subject. By straightforwardly reproducing results, readers will directly test the advantage of a technique with respect to a different one. Finally, promoting RR will make it easier to discover possible false (or inaccurate) results and help the IEEE to maintain its reputation as a world-class scientific/professional organization.

Q: This special issue on reproducibility and measurability of robotics research demonstrated a high interest from the community of the IEEE Robotics and Automation Society (RAS). What opportunities do you foresee for linking this interest to future initiatives in this area which could be launched at the IEEE level?

GS: Developing the necessary infrastructure for supporting RR as well as the best practices associated with it (e.g., in terms of the review process of the data, code, and algorithms associated with the paper) will require substantial work and support from many different IEEE communities. The RAS has already made significant steps in these directions and its experience will be truly precious for the entire organization.

Q: Do you consider research in the robotics and automation domains a key area for applying the principles, methods, and tools for aiming at RR?

GS: Absolutely. Robotics and automation is one of the best areas to apply and test any best practices that the IEEE will develop in terms of RR. In fact, the robotics and automation domains rely

on mathematical science as well as computational and experimental ones, so that these experiences pertain to all kinds of reproducibility mentioned in the answer to the second question.

Q: Do you consider these topics important for the training of a new generation of researchers in engineering?

GS: I consider them fundamental. Adopting RR will, in fact, truly change the culture and will require substantial additional effort from the authors publishing with the IEEE. This is, of course, a process that cannot be enforced but only reinforced. We need, therefore, to educate the community, especially the young professionals, to comprehend and embrace the benefits that RR can bring from all different perspectives: authors (visibility increase), users (enhancement in the exploitability of results, increase in capability of recognizing fundamental results) and humanity as a whole (increased rate of innovation).

References

[1] Wikipedia. Reproducibility. [Online]. Available: <http://en.wikipedia.org/wiki/reproducibility>

[2] J. B. Buckheit and D. L. Donoho, "WaveLab and reproducible research," Dept. of Statistics, Stanford Univ., Stanford, CA, Tech. Rep. 474, 1995.

[3] H. W. Turnbull, Ed., *The Correspondence of Isaac Newton: 1661–1675*, vol. 1. London, U.K.: Royal Society at Univ. Press, 1959, p. 416.

[4] Special online collection: Hwang et al. controversy—Committee report, response, and background. *Science*. [Online]. Available: <http://www.sciencemag.org/site/feature/misc/webfeat/hwang2005/>

[5] C. G. Begley and L. M. Ellis. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*. [Online]. 483(7391), pp. 531–533. Available: <http://www.nature.com/nature/journal/v483/n7391/full/483531a.html>

[6] J. P. A. Ioannidis et al. (2009). Repeatability of published microarray gene expression analyses. *Nature Genet.* [Online]. 41(2), pp. 149–155. Available: <http://www.nature.com/ng/journal/v41/n2/full/ng.295.html>

[7] How science goes wrong. (2013, 19 Oct.). *The Economist*. [Online]. Available: <http://www.economist.com/printedition/2013-10-19>

[8] H. A. Piwowar, R. S. Day, and D. B. Fridsma, "Sharing detailed research data is associated with increased citation rate," *PLoS One*, vol. 2, no. 3, e308, 2007, DOI: 10.1371/journal.pone.0000308.



From the Field (continued from p. 188)

Importantly, the system also allows researchers to create their own specialized, operational tasks to run in natural settings. We claim that this method can be applied to all kinds of applications.

An opportunity to try this out arose in July 2014 at the Arkaroola Mars Robot Challenge organized by the Mars Society Australia. Four student teams brought six field robots to a test site in Arkaroola, a remote desert station in South Australia. The machines embodied the students' design concepts for assistant robots for astronauts performing tasks on the Martian surface (Fig-

ure 1). A selection of six standard DHC–NIST–ASTM benchmarks, together with three operational tests specific to surface operations in harsh Mars-like terrain, was conducted over 12 days. For example, we had the robots search a gullied slope for a hidden target object, which had to be photographed, collected, and returned to the operator.

The test details and results will be formally presented in September at Towards Autonomous Robotic Systems 2015 in Liverpool, United Kingdom, but in brief, we found that most, but

not all, tests worked well, provided one practices the procedures and allows enough time (Figure 2). We were not only able to gather a good deal of standard performance data, but we were able to use it later to make real design improvements to two of the robots. Our test program could accommodate unmanned aerial vehicles as well as ground machines: one participant was able to score highly on many tasks using a small quadrotor, suggesting very high utility of a (suitably modified) drone for future Mars explorers.

