Francesco Amigoni and Viola Schiaffonati

Robotic competitions as experiments: From play to work

Abstracts

In this paper, we plan to investigate the relationship between robotic competitions and scientific experiments. Traditionally robotic competitions have been conceived as useful scenarios for testing robots' abilities in playful settings, given that the competitive element involved makes them attractive to an audience larger than that of experts. More recently, the attempt to frame those competitions as scientific experiments has gained increasing attention. Amongst the various reasons for this reframing, there is the effort to exploit already established infrastructures while developing a more rigorous and systematic approach to the experimental evaluation of robotic artefacts. However, the assimilation of robotic competitions to experiments is not as immediate as it may appear. Building on a careful analysis of results of some major robotic competitions (RoboCup and DARPA Robotic Challenge) we claim that in this context the classical notion of experiment should be reconsidered in the direction of an investigation rather than of a controlled process. The transition from robotic competitions as purely playful events (although with an exceptional amount of technology involved) to kinds of scientific experiments, where the abilities of robots are systematically tested, opens new and interesting ways of considering the interplay between play and work in a technoscientific context.

In diesem Artikel möchten wir das Verhältnis zwischen Roboterwettbewerben und wissenschaftlichen Experimenten untersuchen. Traditionell wurden Roboterwettbewerbe als nützliche Szenarios aufgefasst, um die Fähigkeiten der Roboter in spielerischen Umgebungen zu testen, da das wettbewerbliche Element sie für ein größeres Publikum jenseits der Expertinnen und Experten attraktiver macht. Seit einiger Zeit erfährt der Versuch, diese Wettbewerbe als wissenschaftliche Experimente zu gestalten, zunehmende Aufmerksamkeit. Zu den vielen Gründen für diese Herangehensweise gehört die Anstrengung, bereits etablierte Infrastrukturen zu nutzen, während ein rigoroserer und systematischerer Ansatz entwickelt wird, um die experimentelle Bewertung der Roboter-Artefakte vorzunehmen. Dennoch ist die Anpassung von Roboterwettbewerben an Experimente weniger unmittelbar als es zunächst scheinen mag. Aufbauend auf die gründliche Analyse von Ergebnisse einiger wichtiger Roboterwettbewerbe (RoboCup und DARPA Robotic Challenge) behaupten wir, dass in diesem Rahmen der klassische Begriff des Experiments nochmals dahingehend überdacht werden muss, ob er nicht eher eine explorative Untersuchung als einen kontrollierten Prozess darstellt. Der Übergang von Roboterwettbewerben von rein spielerischen Veranstaltungen (obgleich dabei eine außergewöhnliche Menge an Technik zum Einsatz kommt) zu einer Art von wissenschaftlichen Experimenten, bei denen die Fähigkeiten des Roboters systematisch getestet werden, eröffnet neue und interessante Wege, das Wechselverhältnis von Spiel und Arbeit in einem technowissenschaftlichen Kontext neu zu betrachten.

Introduction

In this paper, we investigate the relationship between robotic competitions and scientific experiments as forms of play (although with some peculiar features) and work, respectively. Rather than considering them as opposite, we aim to show how the currently observed transition of robotic competitions from mostly playful events to contexts for soundly scientific experimentation must be accompanied by a different notion of experiment that we label as *explorative*.

Traditionally, robotic competitions have been conceived as useful scenarios for displaying a robot's abilities in a playful setting, given that the competitive element involved makes them attractive to an audience larger than that of experts. More recently, the attempt to frame those competitions as scientific experiments has gained increasing attention.[1] Amongst the various reasons for this reframing, there is the effort to exploit already established infrastructures, while developing a more rigorous and systematic approach to the experimental evaluation of robotic artefacts. For example, the fact that robotic competitions are repeated regularly over time is a way of promoting repeatability - one of the golden standards of experimental methodology. However, the assimilation of robotic competitions to experiments is not as immediate as it may appear. To argue in this direction, we need to elucidate the recent attention to good experimental methodologies promoted in the autonomous robotic community and to clarify the way in which the notion of experiment is meant in this context.

The discussion on experiments and the effort in developing good experimental methodologies have gained attention in autonomous robotics in the very last years. This field is oriented to develop robot systems that are autonomous in the sense that they have the ability to operate without continuous human intervention,[2] to work in places hardly accessible by humans or in cooperation with humans in common environments. In autonomous robotics, human operators evolve from being active controllers of the robot systems to being more passive supervisors of the same robot systems. There are different reasons for the recent interest in experimentation in the field: from a scientific perspective, it concerns the desire of this rather novel community to adopt the same methodological standards of other scientific disciplines; from a more practical and industrial perspective, it deals with the possibility of measuring some parameters (e.g., safety of a home assistant robot) in a standard way and of having rigorous benchmarks to compare and evaluate different products.

1 See John Anderson, Jacky Baltes and Chi Tai Cheng: »Robotics competitions as benchmarks for AI research,« *The Knowledge Engineering Review* 26 (2011), pp. 11–17.

2 Generally speaking, and for the purpose of our presentation, a robot system is an artefact that interacts with the external environment through its sensors and actuators and that is controlled by software programs.

The interest in solidly-based experimental research in autonomous robotics increased progressively during the last 20 years. It is only recently that this interest has been coupled with a careful analysis on how the concept of experimentation should be translated in the practice of autonomous robotics, also giving rise to a debate about the status of the discipline itself. To this end, both the creation of the EURON Special Interest Group on Good Experimental Methodology in Robotics Research and the series of workshops about replicable experiments in robotics have played a decisive role.[3] In particular, the workshop series has contributed to raise several issues and to increase the sensibility of the community on these topics.[4]

In the effort of improving the quality of experimental activities, some attempts have been made to take inspirations from how experiments are performed in the natural sciences,[5] by trying to translate their general experimental principles (e.g., comparison, repeatability, reproducibility, justification, explanation, …) into the practice of autonomous robotics. However, a recent analysis suggests that these principles are not yet fully part of the current research practice.[6] Notwithstanding the emphasis on the importance of reproducibility in order to increase the experimental level of the field and good practices to promote it, such as the availability of shared data and code, they are still not very common. In addition, the attempts to critically analyze how the main principles should be achieved in experiments with autonomous robots are largely unexplored.

It is true that, dealing with technical artefacts, robotics cannot be plainly assimilated to a traditional scientific field where experiments are generally conducted for hypotheses testing purposes and with a strong theoretical background. At the same time, the type of experiment to which autonomous robotics aspires to conform is a *controlled experiment*, namely:

> »[…] a procedure in which some object of study is subjected to interventions (manipulations) that aim at obtaining a predictable outcome or at least predictable aspects of the outcome. Predictability of the outcome, usually expressed as repeatability of the experiment, is an essential component of the definition. Experiments provide us with informa-

---

3 See Fabio Bonsignorio, John Hallam and Angel del Pobil: »GEM Guidelines«, http://www.heronrobots.com/EuronGEMSig/downloads/GemSigGuidelinesBeta.pdf (visited: 29.01.2017); Fabio Bonsignorio, John Hallam and Angel del Pobil: »Special interest group on good experimental methodologies«, http://www.heronrobots.com/EuronGEMSig/gem-sig-events (visited: 29.01.2017).

4 See Fabio Bonsignorio and Angel del Pobil: »Toward Replicable and Measurable Robotics Research«, *IEEE Robotics and Automation Magazine* September 2015, pp. 32–35.

5 See Francesco Amigoni, Monica Reggiani and Viola Schiaffonati: »An insightful comparison between experiments in mobile robotics and in science«, *Autonomous Robots* 27 (2009), pp. 313–325.

6 See Francesco Amigoni, Viola Schiaffonati and Mario Verdicchio: »Good experimental methodologies for autonomous robotics. From theory to practice«, in: Francesco Amigoni and Viola Schiaffonati, eds., *Methods and Experimental Techniques in Computer Engineering*, Cham: Springer 2014, pp. 37–53.

tion about regularities, and without predictability or repeatability we do not have evidence of anything regular«.[7]

Yet, the reality of autonomous robotics practice is rather different from that of the natural sciences, as robot systems are human-made artefacts. Accordingly, experiments have the goal to demonstrate that a given artefact is working with respect to a reference model (e.g., its design requirements or its expected behavior) and, possibly, that it works better than other similar artefacts with respect to some metrics. This kind of experiment brings them closer to typical engineering tests. To put it simply, experiments in engineering fields have other objects (technical artefacts rather than natural phenomena) and other purposes (testing rather than understanding) with respect to experiments in the sciences. At the same time, the most advanced robot systems are extremely complex, and their behavior is hardly predictable, even by their own designers, especially when considering their interactions with the natural world that are difficult, if not impossible, to model in a fully satisfactory way. In this sense, experiments in autonomous robotics also have the goal of understanding how these complex systems work and interact with the world and, therefore, are similar to natural science experiments.

Within this framework, the assimilation of robotic competitions to experiments is not as immediate as it may appear. We claim that in this context the classical notion of experiment should be reconsidered to be an *investigation* rather than a controlled process. The transition from robotic competitions as purely playful events (although with an exceptional amount of technology involved) to kinds of scientific experiments, where the abilities of robots are systematically tested, is in line with a different view of experimentation, mostly focused on exploration. Here, the emphasis shifts from the traditional conception of experimental control to a different conception of control (*a posteriori*) and to forms of deliberate learning that characterize explorative experiments. Overall, the settings presented are a hopefully promising attempt of considering the interplay between play and work in a technoscientific context.

In this paper, after discussing the evolution of robotic competitions from games to experiments (Section 2), we reason that robotic competitions cannot be considered controlled experiments in a traditional way (Section 3). Then, we attempt to broaden the framework of analysis by borrowing some ideas from the recent debate in the analytical philosophy of technology and in the philosophy of technoscience (Section 4). This will lay the groundwork for the notion of explorative experiments (Section 5). This, in return, is used to reframe robotic competitions (Section 6). Finally, we advance some concluding remarks (Section 7).

7 See Sven Ove Hansson: »Experiments Before Science. What Science Learned from Technological Experiments«, in: Sven Ove Hansson, ed., *The Role of Technology in Science*, Dordrecht: Springer, 2015, pp. 81–110, here p. 88.

Robotic competitions and challenges have flourished since the 1970 s.[8] Now there are dozens of events per year. From the beginning, the ability of competitions playing several roles in robotics was recognized. For instance, competitions have the potential to promote education and research, push the field forward, entertain a general audience, and build a community.[9] These roles often conflict with one another. Balancing them in devising a robotics competition has therefore been proven to be difficult.[10] A robotics competition usually involves some robots, a dynamic, but rather controlled, environment, clear measures of success, and rules for calculating scores and ranking participants.[11]

One of the best-known examples is RoboCup,[12] which has been taking place since 1996 and aims to provide standard problems to ensure the evaluability of various theories, algorithms, and architectures. RoboCup is a very attractive event, both for practitioners and for public in general. For example, when RoboCup 2016 took place in Leipzig (Germany), 3,500 participants were involved from more than 45 countries, who brought more than 1,200 robots to the event and more than 30,000 visitors. RoboCup features soccer, rescue, home, and industrial competitions, in which robots compete in dynamic unpredictable environments with real-time constraints. Competitions take place in both the real world and in a simulation. The environments are precisely defined and can be easily reproduced in different places (and they actually are, in regional RoboCup events). However, this is not true for other elements that characterize the competition, such as opponent teams as well as

---

8 In this paper, we consider robotic competitions and challenges as synonymous and, for the sake of brevity, we mostly speak of competitions.

9 R. Peter Bonasso and Thomas Dean: »A Retrospective of the AAAI Robot Competitions«, *AI Magazine* 18 (1997), pp. 11–23; Robin R. Murphy: »Using Robot Competitions to Promote Intellectual Development«, *AI Magazine* 21 (2000), pp. 77–90; Jennifer Casper, Mark Micire, Jeff Hyams, and R. Murphy: »A Case Study of How Mobile Robot Competitions Promote Future Research«, in: Andreas Birk, Silvia Coradeschi and Satoshi Tadokoro, eds., *RoboCup 2001. LNAI 2377*, Berlin: Springer 2002, pp. 123–132 and Moi-Tin Chew, Serge Demidenko, Chris Messom and Gourab Sen Gupta: »Robotics Competitions in Engineering Education«, *Proceedings of the 4th International Conference on Autonomous Robots and Agents (ICARA 2009)*, Wellington: IEEE, 2009, pp. 624–627.

10 Anderson, Baltes and Cheng: »Robotics competitions as benchmarks for AI research«, pp. 11–17.

11 Holly A. Yanco: »Designing Metrics for Comparing the Performance of Robotic Systems in Robot Competitions«, *Workshop on Measuring Performance and Intelligence of Intelligent Systems (PERMIS)*, 2001; Elena R. Messina, Stephen B. Balakirsky and Rajmohan Madhavan: »The Role of Competitions in Advancing Intelligent Systems: A Practioner's Perspective«, *Proceedings of the 9th Workshop on Performance Metrics for Intelligent Systems*, Gaithersburg: National Institute of Standards and Technology, 2009, pp. 105–108 and James Parker, Julio Godoy, William Groves and Maria Gini: »Issues with Methods for Scoring Competitors in RoboCup Rescue«, *AAMAS Workshop on Autonomous Robots and Multirobot Systems*, 2014.

12 RoboCup, http://www.robocup.org (visited: 29.01. 2017).

light and noise conditions. In the soccer competitions, the measures and criteria according to which two robot systems (teams) are compared are clearly defined only for the purposes of the game and basically consist in the number of scored goals per game (like in human soccer). In other leagues, like in RoboCup@Home, some attempts were recently made toward more solid procedures to benchmark and track the progress of robots in performing tasks.[13] For instance, in RoboCup@Home, robots are tested in tasks involving their abilities to navigate, recognize and track people and objects, manipulate objects, and recognize speech and gestures. Scores are given to the performance of these abilities within a task, allowing to observe their evolution over time.

Another successful and famous example of robotic competitions is represented by the DARPA Robotics Challenge (DRC) which consists of tasks related to human assistance in responding to disasters performed by autonomous humanoid-like robots able to operate in hazardous settings.[14]

After some early recommendations about being careful of not confusing a competition with research,[15] a more recent trend is advocating to recast robotic competitions as experiments, because they offer an excellent vehicle for advancing the state of the art and evaluating new algorithms and techniques in the context of a common problem domain.[16] Along the same line, there was a proposition to use competitions as benchmarks, since they provide standardized test beds that are largely independent of the specific settings roboticists usually experience in their laboratories. This allows a direct comparison of approaches for solving a task.[17] A recent publication has further strengthened the relationship between robotic competitions and scientific experiments.[18] While recognizing a difference in the scope between these two activities, the described approach is representative of the current tendency to promote a more rigorous attitude towards the design of robotic competitions. According to the authors, robotic experiments are characterized by the following features: a common testbed (provided through detailed specifications), specific performance metrics (standards of measurement of some desired characteristics), reproducibility and re-

13 Dirk Holz, Luca Iocchi and Tijn L.: »Benchmarking Intelligent Service Robots through Scientific Competitions. The RoboCup@Home Approach«, *AAAI Spring Symposium on Designing Intelligent Robots: Reintegrating AI II*, 2013 pp. 27–32.

14 DARPA Robotics Challenge, http://www.darpa.mil/program/darpa-robotics-challenge (visited: 29.01.2017).

15 Thomas Bräunl: »Research Relevance of Mobile Robot Competitions«, *IEEE Robotics and Automation Magazine* 6 (1999), pp. 32–37.

16 Monica Anderson, Odest Chadwicke Jenkins, and Sarah Osentoski: »Recasting Robotics Challenges as Experiments«, *IEEE Robotics and Automation Magazine* June 2011, pp. 10–11.

17 Sven Behnke: »Robot Competitions. Ideal benchmarks for Robotics Research«, *Proceedings of the IROS Workshop on Benchmarks in Robotics Research*, Beijing, October 2006.

18 Luca Iocchi, Dirk Holz, Javier Ruiz-del-Solar, Komei Sugiura and Tijn van der Zant: »ROBOCUP@HOME. Analysis and results of evolving competitions for domestic and service robots«, *Artificial Intelligence* 229 (2015), pp. 258–281.

peatability that, although not always performed in practice, should be considered as regulative ideals. In contrast, competitions involve the specification of the competition environment, the specification of robot requirements and constraints, specific performance metrics to rank the participants, and information about how competitions are organized and run. Thus, the environment specification represents a common element that, according to the authors, should be further strengthened in the direction of an experimental approach to robotic competitions.

The interest in designing competitions as experiments has also resulted in two FP7 projects funded by the European Union (EU). euRathlon is an outdoor competition for robots involved in emergency-response scenarios with teams of terrestrial, marine, and aerial robots.[19] The settings in which the competitions take place represent mock emergency-response scenarios, including conditions like limited visibility and salty water. Scores consist of measured quantities and subjective evaluations given by a human Judging Team. Data sets recorded during the competition are made publicly available to the community as a valuable tool for benchmarking, testing and comparison.

The other EU-funded project is RoCKIn which addresses domestic (RoCKIn@Home) and industrial (RoCKIn@Work) environments,[20] with a focus on autonomous service and industrial robots, respectively. Explicit emphasis is put on assessing, comparing, and evaluating competing approaches by means of benchmarking procedures and good experimental methods. One of the distinctive features of RoCKIn is its scoring system that is based on the presence of two classes of benchmarks, called *task benchmarks* and *functionality benchmarks*. The first one is devoted to evaluating the performance of integrated robot systems (in tasks like welcoming a visitor), while the second one focuses on the performance of specific subsystems (like object recognition and localization). Additionally, RoCKIn makes data collected during competitions available to the community.

In general, we can say that competitions share some similarities with experiments. Robots usually compete in precisely definite settings and are scored according to precise performance measures, which parallels, at least to some degree, the controlled conditions and measures of experiments. However, competitions often evaluate whole robot systems, while experiments presented in the literature mainly evaluate a single robot ability or sub-system. Moreover, some competitions are designed to be performed just once, meaning that they are conceived as unique events without considering their reproducibility. In fact, they are not usually reproduced, while experiments aim at being reproducible. Therefore, (generally speaking) robotic competitions usually evaluate general abilities of robot systems and push to development of solutions, while experiments evaluate specific hypotheses, explore phe-

19 euRathlon, http://www.eurathlon.eu (visited: 15.02.2015).
20 RoCKIn, http://rockinrobotchallenge.eu (visited: 15.02.2015).

nomena, and share results. Accordingly, the question whether robotic competitions are experiments or not is far from being satisfactorily answered.

Robotic competitions and controlled experiments

In the previous section, while acknowledging some reasons for supporting the reframing of robotic competitions as scientific experiments, we have reported some of the differences already spelled out by Iocchi et al. for the context of autonomous robotics.[21] In this section, we analyze in depth these differences and claim that, not only is the assimilation between robotic competitions and experiments far to come in practice but, at a more fundamental level, some issues prevent this assimilation from a conceptual point of view.

Let us first consider the difference in the *scope*: if controlled experiments aim at demonstrating and measuring the performance of a system or of a component to solve a particular problem, competitions aim at directly comparing different solutions in a predefined testbed. This difference has impacts especially on the definition of the metrics, such as in experiments where they are usually defined with the aim of measuring the performance of a specific component, while in competitions they measure the system's general ability for ranking purposes. The difference in the scope also has an influence on the definition of tasks in a competition, which in return affects how participants develop their solutions to the specified problems, often by endorsing generally suboptimal solutions that work well under restricted conditions. For instance, if the rules of a robotic soccer competition prescribe that the ball is red, then teams can perform well if their robots recognize red blobs in the images captured by onboard cameras, without possessing any general ability of recognizing balls.

Let us focus on reproducibility and repeatability now. These are considered to be two of the fundamental principles of experimental methodology. *Reproducibility* is the possibility to verify, in an independent way, the results of a given experiment. It refers to the fact that other experimenters achieve the same results by starting from the same initial conditions, using the same type of instruments, and adopting the same experimental techniques. To be reproducible, an experiment must be fully documented. *Repeatability* concerns the fact that a single result is not sufficient to ensure the success of an experiment. A successful experiment must be the outcome of a number of trials performed at different times and in different places. These requirements guarantee that results have not been achieved by chance, but that they are systematic. Clearly, both principles are problematic when applied to the case of robotic

21 Iocchi, et al. »ROBOCUP@HOME«, pp. 258–281.

competitions. Reproducing a robotics competition is extremely difficult to obtain since the recreation of the same scenario is not feasible due to limitations in terms of time, space, and cost. One example of this is represented by the difficulty of reproducing the same surrounding environment in competitions that take place outdoors under different weather conditions. Moreover, repeatability of the same competition is usually not considered for organizational purposes, in particular due to its expensiveness in terms of time and for the meager attractiveness of this kind of activity for the audience.

When regarding both principles, it is difficult to devise solutions without affecting the very nature of competitions. Notwithstanding the efforts in devising robotic competitions with a scientific attitude (see the RoCKIn project cited above for example), practical reproducibility of competitions over the years is also prevented by the fact that some evolution in the task selection is required to prevent overfitting and, therefore, to guarantee the success of the competition. Indeed, periodical competitions change the set of the selected tasks to increase the difficulty of the addressed problems and to provide different experimental conditions. For example, in RoboCup@Home, a task is changed by adding a phase that requires a new robot ability by means of increasing the number or the duration of the task activities (e.g., recognizing more people and track them longer), by removing constraints (e.g., people are not assumed to be seated), and by increasing environment clutter and background noise.[22] Although repeatability could be achieved to some extent when provided with an adequate amount of organizational effort and expense, the temporal evolution of the competitions over the years is such a characterizing feature that a solution for implementing reproducibility is hardly conceivable or is conceivable at the expense of losing the attractiveness of competitions.

So far, we have focused only on the difficulty of reproducing the same environmental conditions in different contexts, implicitly assuming the ability to test the same subject, namely the robots. However, robots are subject to change to comply with evolving requirements and technological development.

A technoscientific framework for experiments

We have argued that robotic competitions cannot be fully assimilated to controlled experiments, despite some similarities and the recent effort of the community to promote a more scientifically-based approach. As mentioned in the Introduction, autonomous robotics is situated in between science and engineering where the way in which experimentation is practically oriented is influenced by this position. The

22 Ibid.

characterization of autonomous robotics as an engineering science might probably be beneficial for reconsidering the relationship between robotic competitions and experiments within a wider and more articulated framework.[23] Clearly, due to the peculiar nature of this filed, the traditional principles of controlled experimentation cannot be applied in a straightforward way. Even if the engineering sciences show a continuity with the natural ones, they work within the context of technological applications, with an interest in the development of technological devices, processes, and materials. Accordingly, experiments are mainly carried out to check whether these technical artefacts meet the desired specifications via the technological production of these phenomena, instead of theories.[24] Technological production must be considered broadly in this discussion, including such activities as research, development, design, testing, patenting, maintenance, inspection, and so on, with attention to the everyday practice of engineers. This suggests an agenda for philosophical reflection on engineering that is distinct from the traditional philosophy of technology, mostly focused on the analysis of the social, cultural, and political impact of technologies.[25] Moreover, technical artefacts, as physical objects with a technical function and use, that have been planned, designed and made by human beings,[26] introduce a normative character to the experimental context in the evaluation of artefacts. This is entirely absent in natural sciences experiments. These aspects set the frame for a *technoscientific* perspective, namely that particular perspective that can be meaningfully summarized as an ›engineering way of being in science‹.[27]

The distinction between science from technoscience is a matter not only of terminological clarification, but also of guiding ideals and research orientations that shape practice in significantly different ways. Using the words of a programmatic paper:

---

23 The meaning of engineering science we refer to is Mieke Boon: »Scientific Concepts in the Engineering Sciences: Epistemic Tools for Creating and Intervening with Phenomena«, in: Uljana Feest and Friedrich Steinle, eds., *Scientific Concepts and Investigative Practice*, Berlin: De Gruyter 2012, pp. 219–243. She defines it as follows: »Engineering sciences, which is scientific research in the context of technological applications, is an example of a science in the context of application. Its purpose is scientific research that contributes to the development of technological devices, processes, and materials«.

24 Mieke Boon: »Understanding Scientific Practices: The Role of Robustness Notions«, in: Lena Soler, Emiliano Trizio, Thomas Nickles and Willian Wimsatt, eds., *Characterizing the Robustness of Science after the Practical Turn of the Philosophy of Science*, Dordrecht: Springer, 2012, pp. 289–315.

25 Ibo van de Poel: »Philosophy and Engineering: Setting the Stage«, in: I. van de Poel, and D. E. Goldberg, eds., *Philosophy and Engineering: An Emerging Agenda,* Dordrecht: Springer 2010, pp. 1–11.

26 Pieter Vermaas, Peter Kroes, Ibo van de Poel, Marteen Franssen and Wibo Houkes: *A Philosophy of Technology. From Technical Artefacts to Sociotechnical Systems*, Williston: Morgan & Claypool 2011.

27 Alfred Nordmann: »Science in the context of technology«, in: Martin Carrier, Alfred Nordmann, eds., *Science in the context of application.* Boston studies in the philosophy of science, Dordrecht: Springer 2010, pp. 467–482.

»Both science and technoscience involve an interplay of representing and intervening. Science is defined by its orientation to the epistemic ideal of purification […]. Technoscience is defined by its neglect or abandonment of this work of purification. […] Technoscience is therefore a kind of research where theoretical representation and technical intervention cannot be held apart even in thought«.[28]

As these authors emphasize, it is not always possible to distinguish if an observable effect is a contribution of the researchers or of nature. Both the observed phenomena and the effects appear to be engineered in a way that there is no interest in a work of purification. For instance, in the case of a pharmaceutical therapy it is irrelevant to try to understand the contributions achieved by technology and by nature to assess the effectiveness of the therapy.

Mentioning biomedical research and laboratory experimental practice is useful to introduce an example of a type of experimental process which cannot be easily accommodated within the traditional notion of an experiment such as, for instance, the clinical trial of an analgesic. Here, the outcome looked for is effective pain reduction and the experimental intervention is the treatment that might be administered. According to Hansson,[29] this is a *directly action-guiding experiment* which satisfies two criteria: (a) the outcome looked for consists in the attainment of some desired goal of human action (e.g., pain reduction) and (b) the interventions studied are potential candidates for being performed in a non-experimental setting in order to achieve that goal (e.g., treatment administered in ordinary conditions). This definition of directly action-guiding experiment frames the notion of experiment as a technological form of experimentation driven by practical needs. As Hansson carefully reconstructs, this form of experimentation has a non-academic origin and was already employed in pre-scientific times, playing a major role during the early Renaissance for the development of experimental methodology. The practical attitude of directly-action guiding experiments is contraposed by Hansson to *epistemic experiments*, namely experiments oriented towards the knowledge of the word we live where the desired outcome is the knowledge itself. Historical and philosophical accounts of experimental methods have been principally focused on epistemic experiments, and little attention was devoted to directly action-guiding ones.

The idea that different notions of experimentation can coexist without any radical contraposition has only recently been pointed out in computer engineering, and still represents an exception to the mainstream framework depicted in the previous section for autonomous robotics. As reconstructed in detail in Tedre's book dated in

28 Bernadette Bensaude-Vincent, Sacha Loeve, Alfred Nordmann and Astrid Schwarz: »Matters of Interest. The Object of Research in Science and Technoscience«, *Journal of General Philosophy of Science* 43 (2011), pp. 365–383, here p. 368.

29 See Sven Ove Hansson: »Experiments Before Science. What Science Learned from Technological Experiments«, in: Sven Ove Hansson, ed., *The Role of Technology in Science*, Dordrecht: Springer, 2015, pp. 81–110.

2015,[30] at least five different views of experiments can be recognized in the computer engineering practice. There are *feasibility experiments*, aimed at empirically demonstrating ('demonstration' and 'experiment' are terms commonly used as synonymous in the field) the proper development and working of a technology. There are *trial experiments*, evaluating some aspects of a system using predetermined variables in a laboratory, and *field experiments*, aimed at evaluating these aspects of a system outside the laboratory, in the real world. There are also *comparison experiments* devoted to finding a befitting solution to a given problem by means of comparison. And, finally, there are *controlled experiments*, similar to those of the more traditional notion of experimentation and aimed at achieving generalization and prediction.

Moreover, Tedre's analysis presents evidence that the notion of control, usually an essential component in experimentation in the natural sciences, plays a different role in computer engineering and becomes critical. If, in the natural sciences, it is prescribed that the experimenter should be an outsider of the phenomenon to be explained, it is not clear how a person working in computer engineering, which is aimed at producing computation-based artefacts, could be an outsider with respect to a phenomenon (i.e., an artefact) that (s)he has created.[31] Except from some significant examples, experiments in computer engineering are usually performed by the same person that has created the artefacts involved in the experiments, losing the sort of independence of the experimenter prescribed in the classical experimental protocol.[32]

The crisis of the traditional notion of experimental control was recently evidenced by Peter Kroes as well.[33] His analysis of experiments with new technologies in socio-technical systems discusses the shift away from a control paradigm based on two assumptions: the experimenter is not part of the system on which the experiment is performed and (s)he is in control of the independent variables and of the experimental set-up. Accordingly, the experimenter is able to intervene both by changing these variables (to evaluate their influence on the dependent ones) and by varying the experimental set-up. This traditional control paradigm becomes problematic when considering new technologies as socio-technical systems, such as hybrid systems com-

30 Matti Tedre: *The Science of Computing*, Boca Raton: CRC Press, 2015.

31 Matti Tedre: »Computing as a Science. A Survey of Computing Viewpoints«, *Minds and Machines* 21 (2011), pp. 361–387.

32 Note that, in this work, we focus mainly on the activities in computer engineering which are carried out in universities and research centers, including robotic competitions, and not on industrial applications, where testing and experimentation are performed also by non-specialists, for example customers. However, also in this last case, a large part of initial testing of computer and robot systems deployed in industrial applications is performed by the computer engineers and roboticists who designed the systems.

33 Peter Kroes: »Experiments on Socio-Technical Systems. The Problem of Control«, *Science and Engineering Ethics* 22 (2016), p. 633–645.

posed of natural objects, technical artefacts, human actors, and social entities. A consequent shift in the notions of intervention and control can be observed. The idea of controlling the experimental system from a center of command that is outside the system becomes highly problematic because the distinction between the experimental system and its environment is blurred and the environment is complex due to the co-presence of technical artefacts and natural and social elements. Hence, the intervention on these systems involves not only controlling both technical artefacts and social elements, but also problematizes the drawing of the line between the experimental system and its environment.

Explorative experiments

The scenario previously described suggests that a framework simply adapting traditionally controlled experiments might be inadequate (following for example the one proposed by Mark Staples, who analyzed engineering epistemology in a Popperian framework[34]). This is not only due to the coexistence of different notions of experimentation in computer engineering (and in autonomous robotics as well), but also to a novel type of experimental intervention where the borders between the experimental setting and the general environment are blurred. The possibility of controlling them from the beginning of the experimental process is also meager. This is the reason why in this paper we use a different notion of experimentation, which is constructed on the conception of the directly-action guiding experiment, but is precisely shaped for autonomous robotics to see whether it can be applied to robotic competitions.[35] Our concept, the *explorative experiment,* emphasizes that factual, action-guiding knowledge (and not only epistemic) reflects the peculiar status of autonomous robotics, a discipline that is constantly struggling between a scientific and an engineering characterization. The use of the term ›explorative‹ is, thus, both to mark down its difference with respect to the concept of the ›exploratory‹ experiment in the philosophy of biology and to stress its specificity to computer engineering.

In some recent philosophical research, exploratory experimentation has been used to label those forms of experimentation in science, which are not always guided by theories. One of the first authors to recognize the epistemic importance of exploratory experiments is Steinle, who defines them as driven by the desire to obtain empirical regularities when no well-formed theory or no conceptual framework is avail-

34 Mark Staples: »Critical Rationalism and Engineering: Methodology«, *Synthese* 192 (2015), pp. 337–362.

35 Francesco Amigoni and Viola Schiaffonati: »Explorative Experiments in Autonomous Robotics«, in: Lorenzo Magnani and Claudia Casadio, eds., *Model-Based Reasoning in Science and Technology, Logical, Epistemological, and Cognitive Issues*, Cham: Springer 2016, pp. 585–599.

able.[36] The same term is used with a slightly different meaning in another article published in the same year, but in the context of some early research in protein synthesis,[37] where exploratory experimentation is considered a style of inquiry not guided by theory. These and other similar works are mainly directed against the theory-driven approaches of most of the philosophy of science, in the spirit of experimentation having a life on its own.[38] They aim at showing the epistemic significance of inquiries presenting several detailed case-studies. The idea that »the aim of exploratory experiments is to generate significant findings about phenomena without appealing to a theory about these phenomena for the purpose of focusing experimental attention on a limited range of possible findings«[39] is probably expressing the recent publications trend providing evidence of the exploratory shift observed in the methodology in some parts of biology.[40]

For our purposes, however, this emphasis on theory is out of scope, as it is not even completely clear what a theory in computer engineering is. In our attempt to characterize explorative experiments in computer engineering and in autonomous robotics we are interested,[41] rather, in the appeal to complexity stressed in philosophical literature,[42] where some systems are considered too complicated to be investigated by means of a theory-driven approach. If this appeal to complexity applies to biology, there are good reasons to apply it to autonomous robotics as well, especially considering that the experimentation subject is not just a technological artefact per se, but also an actor in the surrounding physical and social world.

Explorative experiments are experiments that are driven by the desire of investigating the realm of possibilities pertaining to the functioning of an artefact (which, in the case of autonomous robotics, is a computation-based robot system) and to its interaction with the environment in absence of a proper theory or theoretical background. More precisely, explorative experiments are a special kind of directly action-guiding experiments possessing the following features:

---

36 Friedrich Steinle: »Entering New Fields. Exploratory Uses of Experimentation«, *Philosophy of Science* 64 (1997), pp. S65–S67.

37 Richard M. Burian: »Exploratory Experimentation and the Role of Histochemical Techniques in the Work of Jean Brachet, 1938–1952«, *History and Philosophy of the Life Sciences* 19 (1997), pp. 27–45.

38 Ian Hacking: *Representing and Intervening*, New York: Cambridge UP, 1983.

39 C. Kenneth Waters: »The Nature and Context of Exploratory Experimentation«, *History and Philosophy of the Life Sciences* 19 (2007), pp. 275–284, here p. 280.

40 Laura R. Franklin: »Exploratory Experiments«, *Philosophy of Science* 72 (2005), pp. 888–899.

41 Viola Schiaffonati: »Stretching the Traditional Notion of Experiment in Computing. Explorative Experiments«, *Science and Engineering Ethics* 22 (2016), pp. 647–665 and Amigoni and Schiaffonati: »Explorative Experiments in Autonomous Robotics«, p. 585–599.

42 R.M. Burian: »On MicroRNA and the Need for Exploratory Experimentation in Post-Genomic Molecular Biology«, *History and Philosophy of Life Sciences* 29 (2007), pp. 285–311.

- They are devoted to testing technical artefacts, meant as artificial entities purportedly built by humans to fulfill a purpose and, therefore, having a technical function.
- They are focused on iteratively refining the intervention, meant as the union of knowledge and action characterizing experimental practice, and their ultimate purpose is not to test a general theory, but to probe the possibilities and the limits of the intervention.
- They do not force a sharp distinction between designers and experimenters. Instead, the practitioners often become experimenters.
- The control of the experimental factors cannot initially be fully managed. It is partially carried out after an artefact's insertion into its environment (*a posteriori control*).

Let us focus on control, in particular. Control over experimental conditions usually represents a defining feature of experiments. Not only is it necessary for experiments to intervene actively in the material world, producing all kinds of new objects, substances, phenomena, and processes, but:

> »[…] clearly not any kind of intervention in the material world counts as a scientific experiment. Quite generally, one may say that successful experiments require, at least, certain stability and reproducibility, and meeting these requirements presupposes a measure of control of the experimental system and its environment as well as a measure of discipline of the experimenters and the other people involved in realizing the experiment«.[43]

However, taking explorative experiments into considerations, we observe a different conception of control that is in accordance to the fact that these kinds of experiments are performed to gain confidence on the behavior of robot systems. Therefore, achieving control over the experimental factors is only possible as the experiment proceeds without prior full predictability. For instance, the influence of the environment on the robot's behavior is mostly unpredictable before empirical testing, since usually a solid theoretical framework is not available when dealing with very complex technical artefacts. Thus, in explorative experiments control could be intended in *a posteriori* form, in opposition to *a priori* form that usually takes place in traditional experimental contexts. If, in the latter, experimental factors are fully in control of the experimenter in a sort of anticipation of the scenario to be tested, in the former, the possibility of anticipation disappears and the option for control is in the exploration after a robot system has been introduced to its environment.

Another element to take into account discussing explorative experiments is *learning*. In particular, the form of learning taking place during experimental processes that allow experimenters to improve their knowledge and confidence about artefacts

---

43 Hans Radder: »The Philosophy of Scientific Experimentation. A Review«, *Automated Experimentation* 1 (2009), doi:10.1186/1759-4499-1-2.

on the basis of the outcomes of the experiments. Learning - as the deliberate attempt to learn from the results of an experimental process - was recently proposed as an essential element of experimentation within the framework of introducing new technologies in society as social experiments. In particular, *learning-by-experimentation* was defined as a compromise between *leaning-by-anticipation* and *learning-by-doing*:

> »We might now position learning-by-experimentation between learning-by-doing and learning-by-anticipation. It is similar to learning-by-doing in that it takes place during the actual introduction of a technology in society. Still, it is more anticipatory than regular learning-by-doing because it takes place in a research setting with at least the partial aim to learn something. Ideally then, learning-by-experimentation allows for learning things that cannot be learned by anticipation and at the same time is less costly than learning-by-doing«.[44]

Learning-by-experimentation takes place also in autonomous robotics,[45] where different examples of experiments can be labelled as explorative and can be roughly classified as follows:[46]

Investigating the role of parameters. In this case, experiments are explorative because they are used to elucidate the relationship between the values of the parameters and the behavior of robot systems that the designer only has a rough *a priori* idea of.

Confirmation of expectations or hypotheses. In developing robot systems designers usually consider and build upon a set of expectations regarding the behavior of artefacts when inserted in their operating environments. However, building reliable models of the interaction between robots and their environments is not easy, and hypotheses can be only empirically confirmed since they are not based on a solid theoretical ground. In this case, experiments provide a valuable feedback to the design phases.

Getting insights on the behavior of the robot systems. Another form of explorative experimentation is represented by those processes used to acquire knowledge on how robot systems perform their tasks. These experiments provide quantitative results used by the robot designers to inspect, and possibly modify, the design of the internal methods of the robot system, thus providing a more profound knowledge than the one acquired in investigating the role of parameters.

---

44 Ibo van de Poel: »Society as a Laboratory to Experiment with New Technologies«, in: Elen Stokes, Diana Bowman and Arie Rip, eds., *Embedding and Governing New Technologies*, Singapore: Pan Stanford, forthcoming.

45 Note that – although we discuss what experimenters can learn from the outcomes of experiments in this article – a promising extension of this framework could involve also learning processes that take place in autonomous robots and, in particular, the ways in which robots can autonomously modify their behaviour according to the same outcomes.

46 Further details in: Amigoni and Schiaffonati: »Explorative Experiments in Autonomous Robotics«, pp. 585–599.

Assessing the generality of the robot systems. Amongst the different forms of explorative experimentation, the most elaborated one is that used to gain knowledge about the behavior of robot systems in different settings in order to evaluate their generality. A particularly relevant situation is that of robot systems designed for an environment characterized as error-free, which are later experimentally tested in a setting with noisy data or data that the robot system has never experienced before.

Competitions as explorative experiments

So far, we have demonstrated how the concept of explorative experiment gives reason to a part of the experimental practice in autonomous robotics. In this section, we argue that robotic competitions can be interestingly reframed as explorative experiments. This reframing is not only important from a conceptual point of view, but could also be useful from a practical standpoint in devising appropriate guidelines for a more rigorous and scientifically sound approach to robotic competitions.

Robotic competitions seem to exhibit the four features that characterize explorative experiments. Robotic competitions are devoted to testing technical artefacts against each other, specifically the computation-based technical artefacts that are robots. Moreover, robotic competitions, considered from an experimental perspective, do not test general theories, but rather the possibilities and the limits of robotic system intervention in complex environments. For example, RoCKIn was specifically designed to rigorously test autonomous robots using benchmarks. They were tested both in realistic domestic and industrial environments.[47] Also, competitions as experiments are often performed by the same practitioners who have designed the robots or the software program controlling their behaviors. Finally, the notion of control characterizing explorative experiments (namely, the idea that control cannot be fully managed from the beginning but is in part carried out after a technical artefact has been inserted into its environment) affects robotic competitions too. Even if many elements, particularly those regarding the architecture of the competition (e.g., the size and the color of the fields in which robots play soccer), are under control, other elements such as repeatability and reproducibility are still problematic. For instance, there are limitations in controlling all the issues relative to a robot system before it starts playing a soccer game with unknown teammates and adversaries, as

47 Francesco Amigoni, Enrico Bastianelli, Jacob Berghofer, Andrea Bonarini, Giulio Fontana, Nico Hochgeschwender, Luca Iocchi, Gerahrd Kraetzschmar, Pedro Lima, Matteo Matteucci, Pedro Miraldo, Daniele Nardi and Viola Schiaffonati: »Competitions for Benchmarking. Task and Functionality Scoring Complete Performance Assessment«, *Robotics & Automation Magazine* 22/3 (2015), pp. 53–61.

done in RoboCup.[48] Hence, a further development of the *a posteriori* control concept might be important in the robotic competitions context in order to maintain control as a central focus, even if it is in a reduced manner. The DRC (DARPA Robotics Challenge) that was mentioned previously provides some evidence of such *a posteriori* form of control. The competition was designed having (implicitly) humanoid robots in mind, although their use was not explicitly enforced by the rules. It turned out that the winner team, Team KAIST from South Korea, used a humanoid robot able to transform and switch back and forth from a walking biped to a wheeled platform by simply kneeling down. According to many observers, this was the key to its victory, emphasizing the role of architectural flexibility that was not adequately considered in the competition design. Another example could be that of a competition for home robots fulfilling a task like »welcoming a visitor«. The score reflects the number of achieved goals, like identifying the visitor correctly, opening the door, greeting the visitor, and so on. However, as a form of *a posteriori* control, designers of the competition could decide to also include the average distance between the robot and the visitor into the score, because they noticed unexpected behaviours during the runs at the competition.

In the previous section, we have argued that, in the context of explorative experiments, learning, and in particular learning-by-experimentation, acquire further importance as a defining element of experimentation. The emphasis on learning is also an important element for robotic competitions, when their experimental structure is investigated. The deliberate attempt to learn characterizes robotic competitions in a way that we indeed might say that learning is one of the reasons why competitions are devised and carried out. From the participant's point of view, learning is acquiring knowledge about how robots perform in order to improve them and to design better robots. From the organizers' point of view, learning involves evolving the competitions to keep them challenging for the participants and appealing for the public. In this sense, learning-by-experimentation seems to depict the form of learning carried out in robotic competitions properly.

In conclusion, robotic competitions present some of the characteristic elements of explorative experiments that we have listed in the previous section and, for this reason, they are better conceived in the framework of explorative experiments rather than in that of controlled ones.

48 See Katie Genter, Tim Laue and Peter Stone: »Three Years of the RoboCup Standard Platform League Drop-in Player Competition. Creating and Maintaining a Large Scale Ad Hoc Teamwork Robotics Competition«, *Autonomous Agents and Multi-Agent Systems* 31 (2017), pp. 790–820.

Conclusions

In this paper, we have argued that the recent attempt in the autonomous robotics community to equate robotic competitions to controlled experiments presents several limits, not only because of the lack of experimental practice in that field, but also because of some important conceptual differences. We have thus proposed to reconsider the traditional notion of experiment in a technoscientific context and have found traces of a novel type of experimental intervention where the border between experimental setting and environment is blurred, as well as the possibility of controlling the experimental factors from the beginning. Being inspired by the idea of directly action-guiding experiments, we have therefore proposed the novel notion of explorative experiments and demonstrated how this reflects the experimental practice of autonomous robotics. Finally, we have discussed how robotic competitions can be profitable when reframed as explorative experiments. This reframing does not negate the actual trend of seeing robotic competitions as experiments; rather strengthens it further. However, this is restricted to an explorative context that appears to be more appropriate to an engineering science such as autonomous robotics. Moreover, it promotes an active interaction between play and work, two elements that fruitfully coexist and influence each other in robotic competitions. Even if not investigated in this paper, we believe that the idea of experimentation as performance might be an interesting point towards this direction and that guidelines for robotic competitions should be proposed accordingly.[49]

49 Robert P. Crease: *The Play of Nature. Experimentation as Performance*, Bloomington: Indiana UP 1993.