

An insightful comparison between experiments in mobile robotics and in science

Francesco Amigoni · Monica Reggiani ·
Viola Schiaffonati

Received: 10 February 2009 / Accepted: 5 August 2009
© Springer Science+Business Media, LLC 2009

Abstract Experiments are essential ingredients of science, both to confirm/refute a theory and to discover new theories. It is a common belief that experimentation in mobile robotics has not yet reached a level of maturity comparable with that reached in science, for example in physics, considered as the paradigm of a mature, stable, and well-founded scientific discipline. In this paper, starting from a representative sample of the current state of the art, we identify some basic issues of experiments in mobile robot localization and mapping. These issues, when viewed in the context of some general principles about experiments in science and engineering, lead us to derive some insightful considerations on the role of experiments in mobile robotics. Reflecting the background of the authors, the paper has an interdisciplinary nature at the meeting point of mobile robotics and philosophy of science.

Keywords Experiments · Localization · Mapping

1 Introduction

Starting from the Scientific Revolution of the XVII century, natural scientists have developed the modern scientific method, progressing from simple collections of observations to the extensive use of experiments for investigating natural phenomena. Experiments are so fundamental in

science that the ‘scientific method’ coincides with the ‘experimental method’, at least in the natural sciences. Experiments are performed for discovering new hypotheses that explain natural phenomena, for confirming/refuting theories, and for choosing among competing hypotheses. The scientific method, which has proved very successful for natural sciences, could also be useful in engineering, for example when the behavior of a system or its performance are difficult to characterize analytically. Therefore, as we argue in this paper, engineers have the valuable opportunity to learn from the natural sciences, whose experimental methodologies have been developed and optimized along the centuries. At the same time, they should also consider some fundamental differences in experimentation between natural sciences and engineering.

It is a common belief that experimentation in mobile robotics has not yet reached a level of maturity comparable with that of experimentation in science and in other engineering disciplines. Recently, the interest in experimental methodologies increased dramatically within the mobile robotics community, both from researchers and from funding agencies, according to the idea that good experimental activities could reduce the gap between research and industrial applications. Some projects have been funded by the European Community, including Rawseeds (2006) and RoSta (2007), and series of workshops on benchmarks (EURON GEM Sig 2007) and on measures for intelligent and autonomous systems (Permis 2000) have been held in the latest years. Also some papers on these topics started to appear in major conferences (Amigoni et al. 2007).

Despite the ongoing efforts and the recognized importance of experiments for rigorously evaluating new approaches and for reporting them in an objective and complete manner, these ideas have not yet become really part of

F. Amigoni (✉) · V. Schiaffonati
Dipartimento di Elettronica e Informazione, Politecnico di
Milano, Milano, Italy
e-mail: francesco.amigoni@polimi.it

M. Reggiani
Dipartimento di Tecnica e Gestione dei Sistemi Industriali,
Università di Padova, Vicenza, Italy

current practice, as it can be seen by having a look at recently published papers. In our opinion, this is due, among other reasons, to the difficulty to perform time-consuming experiments under the publish-or-perish pressure and to the weak awareness of experiments as fundamental elements in the development of a robotic system.

In this paper, we argue in favor of this last issue, in the context of mobile robot localization and mapping. *Localization* is the process of estimating the pose of a robot in a map, while *mapping* is the process of building a map of an environment, by integrating perceptions taken at different poses. These two processes are tightly correlated and are globally referred to as *SLAM* (*Simultaneous Localization And Mapping*) (Thrun et al. 2005). In the context of this special issue, localization and mapping enable stable navigation solutions, namely the ability of a robotic system to sense and create internal representations of its environment and estimate pose (where pose consists of position and orientation) with respect to a fixed coordinate frame. This provides mobile robots with the ability to identify obstacles and hazards in the environment, and to maintain an estimate of where they are and where they have been.

In this paper, we start by describing some experimental trends that emerged from a survey of a significant sample of papers in mobile robot localization and mapping (Sect. 2). Then, we do an *excursus* over the very idea of experiment, as it has been shaped in science and engineering, with the goal of identifying some principles that define and characterize experiments in general (Sect. 3). Finally, with these principles in mind, we go back to mobile robotics and we discuss some insightful issues on the role of experiments in localization and mapping (Sect. 4). Reflecting the background of the authors, the paper has an interdisciplinary nature at the intersection between mobile robotics and philosophy of science.

The general purpose of this paper is to contribute to the discussion on the definition of an experimental methodology for mobile robotics. In this sense, our scope is broader than the definition of benchmarks (that are anyway basic ingredients of a sound methodology, see Bonsignorio et al. 2007) and includes also a discussion on the purposes of experimental activities. However, we explicitly note that we do not aim at proposing any definite experimental methodology, but only at giving a contribution in the direction of this definition.

2 Trends in experimental activities in mobile robot localization and mapping

In this section, we outline some interesting trends emerging from experimental activities reported in published papers in the area of mobile robot localization and mapping. These

trends will be compared with some general principles about experiments, individuated looking at science and engineering (Sects. 3 and 4). The purpose of this section is to provide a representative picture of how experimental activities are carried out in mobile robot localization and mapping. We do not aim at providing any comprehensive survey of the field. Instead, we analyzed a number of papers, listed in Table 1, that appear relevant to our goal. Some comments are worth about the selection of the papers. We deem that the papers we chose constitute a good sample, since they cover several themes (including classical and more recent SLAM techniques, outdoor SLAM, visual SLAM, scan matching), extend over almost two decades (from early 1990s to present), have been written by authors from different continents, and have been published in highly-respected journal and conferences. Although, like in any selection, there is a subjective bias, we do not consider it to be significant enough to harm the representativeness of the global picture we provide.

For the sake of clarity, we organize the following presentation by discussing the purposes for which experiments are reported in papers, the data sets on which experiments are performed, and the measured quantities that are reported to support claims. Since all these aspects are closely related with each other and a linear presentation cannot account for all the relationships, we try to highlight some of these relationships throughout the text.

2.1 Purposes of experiments

Experiments in mobile robot localization and mapping can be devoted to demonstrate that the proposed system works, to demonstrate that the proposed system works better than other systems, to get insights on the behavior of a system and on its limits of applicability, or, more frequently, to a mixture of these goals. Let us consider these purposes in more detail.

Localization and mapping algorithms are often hard to characterize theoretically, for example through asymptotic analysis of worst-case and average-case behavior. Therefore, only few papers use analysis of computation complexity to define machine-independent bounds on running time (Leonard and Durrant-Whyte 1991). To demonstrate that a system works, usually experiments are performed to evaluate correctness and accuracy of the proposed algorithms. Examples of papers in which experiments are designed with this goal in mind are Davison (2003), Newman et al. (2002, 2006), Paz et al. (2008).

Papers often compare performance of the proposed algorithms with competing algorithms to demonstrate their superiority (and that of the embedded ideas). This could result in what Johnson (2002) calls a *horse race* paper, where results are published for precisely defined, standardized benchmarks to demonstrate the scientific value of an

Table 1 Papers selected to provide a representative picture of experimental activities in mobile robot localization and mapping

| Authors (Year) | Title | Published in |
|----------------------------------|---|--|
| Amigoni (2008) | <i>Experimental evaluation of some exploration strategies for mobile robots</i> | IEEE Int. Conf. on Robotics and Automation |
| Comport et al. (2007) | <i>Accurate quadrifocal tracking for robust 3D visual odometry</i> | IEEE Int. Conf. on Robotics and Automation |
| Davison (2003) | <i>Real-time simultaneous localisation and mapping with a single camera</i> | IEEE Int. Conf. on Computer Vision |
| Davison et al. (2007) | <i>MonoSLAM: Real-time single camera SLAM</i> | IEEE Trans. on Pattern Analysis and Machine Intelligence |
| Frese et al. (2005) | <i>A multilevel relaxation algorithm for simultaneous localization and mapping</i> | IEEE Trans. on Robotics |
| Grisetti et al. (2007) | <i>Improved techniques for grid mapping with rao-blackwellized particle filters</i> | IEEE Trans. on Robotics |
| Gutmann and Konolige (1999) | <i>Incremental mapping of large cyclic environments</i> | IEEE Int. Symp. on Comp. Intelligence in Robotics and Automation |
| Hahnel et al. (2003) | <i>An efficient fastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements</i> | IEEE/RSJ Int. Conf. on Intelligent Robots and Systems |
| Leonard and Durrant-Whyte (1991) | <i>Simultaneous map building and localization for an autonomous mobile robot</i> | IEEE/RSJ Int. Workshop on Intelligent Robots and Systems |
| Liu and Thrun (2003) | <i>Results for outdoor-SLAM using sparse extended information filters</i> | IEEE Int. Conf. on Robotics and Automation |
| Lu and Milios (1997) | <i>Globally consistent range scan alignment for environment mapping</i> | Autonomous Robots |
| Minguez et al. (2006) | <i>Metric-based iterative closest point scan matching for sensor displacement estimation</i> | IEEE Trans. on Robotics |
| Montemerlo et al. (2003) | <i>FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges</i> | Int. Joint Conf. on Artificial Intelligence |
| Neira and Tardos (2001) | <i>Data association in stochastic mapping using the joint compatibility test</i> | IEEE Trans. on Robotics and Automation |
| Newman et al. (2002) | <i>Explore and return: Experimental validation of real-time concurrent mapping and localization</i> | IEEE Int. Conf. on Robotics and Automation |
| Newman et al. (2006) | <i>Outdoor SLAM using visual appearance and laser ranging</i> | IEEE Int. Conf. on Robotics and Automation |
| Paz et al. (2008) | <i>Large-Scale 6-DOF SLAM With Stereo-in-Hand</i> | IEEE Trans. on Robotics |

algorithm. An example of a paper in which experiments are mainly focused on comparison is Montemerlo et al. (2003), where the FastSLAM2.0 algorithm is compared with a previous version of the algorithm (FastSLAM) and with alternative systems (Extended Kalman Filter methods). Other examples are the comparison of some scan matching systems in Minguez et al. (2006), the comparison of Sparse Extended Information Filter with EKF in Liu and Thrun (2003), and the comparison of exploration strategies in Amigoni (2008). A relevant issue is that system comparison is often performed without detailing the tuning of the parameters. As an exception, in Minguez et al. (2006) some comments about this tuning process are reported.

Several papers use experiments to better understand strengths, weaknesses, and operation of algorithms. This ac-

tivity is important because it helps to identify the optimization criteria, the issues that deserve additional investigation and should be optimized, and the influence of parameters. For example, in Grisetti et al. (2007) the behavior of some parameters over time is analyzed to show their effect on the performance. In the same paper, the authors also discuss some situations in which the system does not perform well, i.e., situations in which the scan match fails. Another example is Frese et al. (2005), where the values of some matrices used by the proposed system are studied and discussed.

2.2 Data sets

Meaningful experimental results require the availability of realistic testbeds and of challenging environments in which

algorithms can be studied and systems compared (Hanks et al. 1993).

Researchers sometimes evaluate systems using data sets collected by themselves, often in their laboratory (Davison et al. 2007; Minguez et al. 2006). While these data sets may be useful for demonstrating that the system works, they do not allow comparison with competing systems. The need for facilitating the evaluation and comparison of different algorithms has driven the development of the Robotics Data Set Repository—RADISH (Howard and Roy 2003), which provides a collection of publicly available data sets for localization and mapping. These data sets are basically acquisitions of measurements (e.g., laser range data and odometry data) performed in different environments. With a similar goal, the European project Rawseeds (2006) is working on the development of a comprehensive, high-quality benchmarking tool. One of its goals is to be able to provide data sets that include the corresponding ground truth, time-synced with the sensor data. This is a challenging issue because, so far, the problem of ground truth availability has been only partially solved, for example using simulated data with incorporated random errors (Comport et al. 2007; Liu and Thrun 2003; Lu and Milios 1997).

2.3 Measured quantities

The surveyed papers employ several qualities to evaluate localization and mapping systems. These qualities can be classified in *intrinsic*, when they refer to system efficiency, and *extrinsic*, when they evaluate the quality of the final solution.

The key intrinsic measurement is the *computational time* required to achieve the final solution. For example, in Montemerlo et al. (2003) different systems are compared according to the time required to process the Victoria Park data set (Guivant et al. 2000). Several measurement units for time have been proposed, ranging from seconds (Davison et al. 2007; Minguez et al. 2006; Newman et al. 2006), to flops (Neira and Tardos 2001), to number of iterations (Frese et al. 2005; Minguez et al. 2006), in an effort of using machine-independent measurements. Whenever this is not possible, system characteristics, such as compiler, operating system, processor, memory, and system load, should be included in the description of the experimental evaluation. This information is rarely reported in papers, with few exceptions; for example Grisetti et al. (2007) reports memory usage.

To get some insights on how the proposed system works, several papers (Davison 2003; Davison et al. 2007; Grisetti et al. 2007; Paz et al. 2008) include a *profiling of the total time*, breaking it down to the time required by each single phase of the system's process. Other papers, as already said, propose analysis of computational complexity (Leonard and Durrant-Whyte 1991). Sometimes, the relation between algorithm performance (speed, robustness, accuracy) and some parameters is also presented. For example,

the number of particles is related to performance in Grisetti et al. (2007), Montemerlo et al. (2003), with the idea that the smaller that number, the faster the system.

In Neira and Tardos (2001), the *robustness*, measured as the fraction of correct hypotheses returned by their system for data association in mapping, is analyzed against the error on the position of the robot. In a similar way, in Minguez et al. (2006), robustness of a scan matching system is evaluated according to the number of correct alignments it produces.

The result of localization and mapping algorithms is exploited by mobile robots to support other processes (e.g., navigation). It is therefore important to experimentally evaluate the extrinsic qualities of these algorithms, namely their result.

The first quality we discuss is *accuracy*, measured by the distance (according to some metric) between the output of the system and the ground truth. For example, in Leonard and Durrant-Whyte (1991), the locations of features estimated by the system are compared with hand-measured locations of the same features. In other cases, ground truth is obtained with GPS measurements (Montemerlo et al. 2003), with theodolites (Neira and Tardos 2001), or with plumb-line (Davison et al. 2007). Several metrics have been proposed to calculate the distance from ground truth, including RMS (Montemerlo et al. 2003), error in the pose (Davison et al. 2007; Lu and Milios 1997), and grid-based metrics (Newman et al. 2006).

Another quality which is often considered is the *topological correctness* of produced maps, usually evaluated by visual inspection. Generally, visual inspection consists in checking whether the resulting maps are topologically consistent (Frese et al. 2005; Minguez et al. 2006) or evaluating the reconstructed path followed by the robot (Lu and Milios 1997). In some other cases, visual inspection evaluates the correspondence between the produced map and a (somehow imprecise) ground truth, like aerial images (Newman et al. 2006) and Google Earth maps (Paz et al. 2008). To check the correctness of the generated maps, they are often magnified while searching for inconsistencies (Grisetti et al. 2007). In the search for impartiality, the same paper argues for “blind” evaluation, namely, evaluation by a non-author. Finally, most of the surveyed papers use loop closure tests, a challenging issue in SLAM, to demonstrate the correctness of the produced maps (Gutmann and Konolige 1999; Hahnel et al. 2003).

A final extrinsic quality is *usefulness of a map for navigation*. An interesting approach is presented in Newman et al. (2002), where the difference between the initial position of a robot and the position of the robot when returning back to the initial position after exploration is measured. In this way, the usefulness of the map for returning to the initial position (i.e., for a path planning task) is evaluated.

A characteristic of experiments in mobile robotics is that it is not always possible to test a single component isolated

from the others. For example, a mapping module is usually tested as a component of a complete robotic system for mapping unknown environments. In this case, some other components (like the module used for scan matching) can influence the final result, as pointed out in Frese et al. (2005).

3 Experiments in science and engineering

In this section we discuss the concept of experiment, the way it is intended in science (and in physics in particular), its main features, and how it has evolved in the last centuries. We start from some historical remarks: we analyze how this concept has been developed from the birth of modern science to its specialization in different disciplines and how it is currently defined in engineering disciplines. Finally, we isolate and discuss some of the “principles” that constitute the core of the scientific experimental methodology.

3.1 The roots of experimental methodology

Our modern conception of experiment has emerged during the Scientific Revolution of the XVII century. Since then, science has become experimental, where ‘experimental’ means based on experiments, which are more than simple collections of observations. An experiment can be seen as a *controlled experience*, namely as a set of observations and actions, performed in a controlled context, to support a given hypothesis. In general, while experiments are performed in controlled conditions, this does not hold for observations. For example, observing a drop of water through a microscope does not constitute an experiment. On the contrary, observing the same drop, after having colored it with a chemical reagent in order to evidence some microorganisms, can be considered an experimental procedure performed to test the behavior of the drop under some controlled circumstances.

Two issues seem central to grasp the very idea of experiment: the possibility of *controlling* some of the features of a phenomenon under investigation and the *purpose* for which an experiment is performed. Control deals with the idea that experiments consist in producing controlled circumstances, in which it can be assumed that a studied relationship does not depend on changes in non-controlled factors, either because they are constant or because the relationship under investigation has been made independent of the variations of these uncontrolled factors. In other words, in an experiment the phenomenon under investigation must be treated as an isolated object. This is clearly an artificial situation: it is assumed that other factors, which are not under investigation, do not influence the studied relationship. This is the reason for experiments being performed in the artificial conditions of laboratories, since in the real world it is not usually possible to control every factor that can influence the studied

phenomenon. As regarding their purposes, experiments can play many roles in science. Experiments are made to test theories and to choose between rival hypotheses or beliefs; moreover, they are used to verify or falsify hypotheses and to concretely apply the theories that have been tested; experiments can improve instruments and our abilities to manipulate objects; in a word, they are able to enhance scientific knowledge and let it progress.

There exists a long tradition (starting from Francis Bacon) that considers an experiment as a sort of question, posed in a specific situation and dealing with some parameters measured in the course of the experiment itself. Let us consider for example the experiment that allowed Galileo Galilei to discover the famous law of falling bodies, claiming that the distance travelled by a falling body is directly proportional to the square of the time it takes to fall. In this case the controlled parameters chosen by Galileo were the space covered by a falling body and the time employed to cover this space, in an ideal situation of absence of air (vacuum). The question associated to this experimental situation is the following: does a mathematical constant relation between the values of these quantities hold? The answer given by Galileo is not only that this constancy exists, but also that the covered space is proportional to the square of the time employed to cover it. Therefore, this experiment achieves a universal result, which depends on the parameters chosen and measured in the experiment itself. The choice of the experimental factors and the possibility to control them, by isolating them from all the other factors that are considered not relevant, are central for any successful experiment.

Despite these general characterizations, a precise definition of experiment is not straightforward. Rather than a definition, a non-exhaustive list of properties can be useful to better shape this concept. Experiments make use of precise measurements in order to quantitatively describe the phenomenon under investigation; they must be repeatable at different times and in different places to check the validity and universality of their results; they must be reproducible by other scientists to confirm that their results are independent of the precise details of the experiments themselves; they must be comparable in order to compare results of different experiments; they must be described with a precise language to give rigor and precision to experimental data; they should use measurement instruments, when possible, to enhance human capabilities.

These very general features of experiments are in accordance with the view of science originated at the beginning of Scientific Revolution: science is basically physics, considered as a unique *corpus* without any further differentiation within it. Accordingly, the philosophical and epistemological issues that arise in experimental practice reflect this view and are devoted to general problems, such as to investigate the general reasons for believing in experimental results.

Today, both the characterization of experiments and the analysis of the correlated philosophical problems reflect the high specialization of science. Since the XVII century, science has progressively become more specialized and, today, it is almost impossible to reflect on science in its generality. Today's science is composed of a large number of specific disciplines that range from physics to economics and psychology, that were not even considered as scientific few decades ago. Within each one of such disciplines the general features of experiments need to be concretely translated and the corresponding philosophy of experiments has to deal with specific experimental problems that can largely differ across disciplines. In a sort of counterposition to the unity of science of modern times, we can now observe an exasperate fragmentation. In this context, trying to define what is an experiment and what are its purposes appears at least naïve. Accordingly, philosophers of experiments do not discuss experimental problems in general physics, but in specific fields of physics, such as quantum mechanics and, in some cases, in even more specific subfields, such as relational quantum mechanics.

3.2 Experiments from science to engineering

If there is a deep connection between the view of experiments in science and their views in specific scientific disciplines, there is instead a gap if we consider the relationship between experiments in scientific disciplines and in engineering disciplines, such as computer engineering. Within engineering disciplines, experiments are regarded in a more pragmatic way, thus moving away from the general features that have characterized them from the Scientific Revolution and that have been inherited in current scientific disciplines, such as physics. In engineering, emphasis has been put on concrete results and most work has been devoted to set out strategies for a good experimental design practice (Montgomery 2005). Experimental design is, from the one side, still tied to the general experimental attitude promoted by science but, from the other side, it tries to adapt some general experimental features to the building of concrete artifacts. For example, according to Barr et al. (1995), experimentation is a process composed of well-defined steps starting from the definition of the goal of the experiment, passing through the choice of the factors to explore, to the design and execution of the experiment, to the analysis of the data and the drawing of conclusions, and ending with the reporting of the results. In this context, an experimental methodology appears to be a list of strategies and well-organized resources that can be exploited whenever it is necessary (see also Johnson 2002 for another example of such approach in computer engineering).

This more pragmatic attitude is probably due to the different purposes experiments have in physics and in computer engineering. As we have seen, experiments in physics

are performed to test theories and to provide the ground for scientific knowledge. They can call for new theories, either by showing that an accepted theory is not valid or by showing new phenomena that need to be explained. They can also provide evidence for the existence of entities involved in theories or that could be explained by a future theory (Franklin 2009). On the other hand, the purposes of experiments in computer engineering and, specifically, in algorithms and heuristic methods are more pragmatic. Doing experiments with algorithms seems aiming at learning something about their correctness and performance. Experiments are thus conducted for specific purposes: for example, to demonstrate a known truth, or to check the validity of an hypothesis, or to examine the performance of a new system. In this context, theories are relegated on the background and seem to play just the role of a general framework in which experiments are conducted: usually an experiment is not made to confirm a general theory, but to check if a system works appropriately.

Experiments in mobile robot localization and mapping are in line with the above engineering perspective. However, in this field a stable experimental methodology, even in the engineering sense of a set of strategies for a good design and analysis of experiments, is still lacking. Researchers do not refer to any common methodology and, more often than not, adopt a do-it-yourself approach. This is perhaps due to the empirical difficulties of robotics: differently from other fields of computer engineering, such as algorithmics, robotics has to deal with the real world and, hence, with the related difficult problems. Given this state of affairs, in the last years several efforts to define experimental methodologies for mobile robotics have been done, as we have already pointed out. In this work, we propose some ideas for contributing to this discussion. We strongly believe that both the two perspectives on experiments—the scientific one and the engineering one—must be taken into account. On the one hand, it can be very useful to draw inspiration from other fields of computer engineering, such as algorithmics, that have already developed a partially mature discussion on some methodological problems of experiments. On the other hand, we want to propose the idea that an experimental methodology is more than a set of procedures that, once adopted, lead in a secure way to the expected result. For this reason, we propose to consider some of the general features of experiments in science, and in particular in physics, that can inspire insightful considerations on the role of experiments in localization and mapping.

3.3 Principles of an experimental methodology

Having claimed for their utility, we now discuss some experimental principles holding in science, and in physics in particular, that can be worth considering also in the experimental procedures of localization and mapping. These principles

are: comparison, reproducibility/repeatability, and justification/explanation.

They constitute the very core of the modern conception of experiments, in which they are so deeply rooted that they are not even made explicit anymore in well-developed experimental disciplines. They represent some defining characteristics of experiments that, in our opinion, need to be further elaborated in the context of mobile robot localization and mapping. They were developed in the context of Modern Scientific Revolution in the XVII century and in the subsequent view of science as a collective activity. From there on, science became the activity of groups of strongly connected scholars, widespread all over Europe, and highly interested in exchanging their results and achievements. The ancient conception of a single scientist (or, better, of a philosopher of nature, as scientists were called until XIX century), capable of carrying out his/her whole research in isolation, slowly faded off, due both to the exceptional growth of scientific knowledge, almost impossible to be managed by a single person, and to the acknowledgment of science as a critical activity (Westfall 1971). This latter point is central in the development of the conception of experiment considered as more than just a collection of observations. Better achievements can be gained if knowledge is at disposal and can be exchanged among scholars. Clearly, the possibility to compare different results and to make reproducible insightful experiments is at the basis of this new attitude.

Let us consider in more detail the principles.

(1) Comparison

At a higher level, comparison means to know what has been already done in the past within the same field of research, both for avoiding to repeat uninteresting experiments and for getting suggestions on what the interesting questions could be. At a lower level, comparison refers to the possibility for future researchers to accurately compare their new results with the old ones. If these features are easily given for granted in principle, serious difficulties can arise in actual comparisons. In particular, a direct comparison may be problematic. To be comparable with others, an experiment must be thoroughly documented. Let us consider the example of a comparison between algorithms: although the algorithms may be the same, they can be implemented differently and run on different test instances, or their running times cannot be referred to the same machine/operating system (Johnson 2002). Moreover, comparison should be accomplished on the basis of a “sincerity” principle that often can be easily disregarded. One should report any strange or unexpected phenomenon encountered during experimentation. This is not just for reasons of intellectual honesty, but also because anomalies can reveal something important and bring to new discoveries. The discovery of the planet Neptune in 1846 is a well-known example of how an anomaly

can turn out to be decisive for a new discovery. According to the Newtonian theory, astronomers were able to calculate the hypothetical orbit of Uranus, at that time considered the most distant planet of the Solar system. However, this hypothetical orbit was not in accordance with the orbit actually reported in well-documented series of observations. This incongruence was not considered sufficient to reject the whole Newtonian theory, but gave some hints on the fact that something in the theory itself needed a revision. Two astronomers, Adams and Le Verrier, tried to explain this anomaly by postulating the existence of a further planet not yet observed. They first assumed the existence of a new planet (Neptune), farther than Uranus; they calculated its mass and position; and, eventually, they observed it in a position almost identical to the calculated one (Grosser 1962).

(2) Reproducibility and repeatability

These features are often confused but, although tightly connected, they refer to different desiderata. They are also related to comparison and to the very general idea that scientific results should undergo to the most severe criticisms in order to be strongly confirmed.

Reproducibility is the possibility to verify, in an independent way, the results of a given experiment. It refers to the fact that other experimenters, different from the one claiming for the validity of some results, are able to achieve the same results, by starting from the same initial conditions, using the same type of instruments, and adopting the same experimental techniques. As in the case of comparison, 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 the result has not been achieved by chance, but is systematic. Repeatability should be intended here in a wide meaning as, according to Hacking (1983), real repeatability is never realized in practice. Typically, repetitions of an experiment are attempts to do the same thing better, namely to produce a more stable, less noisy version of the phenomenon created during the experiment itself. Even when the goal is to try to make precise measurements, what is called for is a better experiment, that increases the precision of measurements so that systematic errors can be eliminated. The only cases of literal repetitions of experiments are those in which people do not believe experimental results, and repetition is made to overcome this skepticism.

A particularly clear example to illustrate both reproducibility and repeatability is given by the controversy about the supposed discovery of the cold fusion claimed in 1989 by Stanley Pons and Martin Fleischmann, two chemists at the University of Utah and at the Southampton University,

respectively (Cromer 1993). In a news press, they claimed to have produced nuclear fusion in a tabletop experiment, reporting anomalous heat production of a magnitude that (they asserted) would defy any explanation, except in terms of nuclear processes. They interpreted the absence of neutrons in their experiment as the proof for a new type of nuclear reaction. The paper describing this supposedly revolutionary result (it was rejected by *Nature*, but accepted for publication in the *Journal of Electroanalytical Chemistry*) contained just five references; three of them were to their precedent works, without any reference to the vast literature on the problem in the nuclear physics field, thus revealing a complete ignorance of the previous work. Moreover, they did not give sufficient details to reproduce their experiment nor adequate proofs that their results were the effects of systematic trials. After some time, it turned out that they forgot some basic procedures in their experimentation and that the extraordinary results they claimed to have achieved were not repeatable. In the news press, which unusually took place before the publication of their paper, they declared to have worked on cold fusion in secret during the preceding five years. Also this particular shows that their behavior was not experimentally nor scientifically sound. Science as experimental activity cannot be conducted in isolation. In the achievement of experimental results, other scientists' comments and critiques have a fundamental role to revise work and to check its validity.

(3) Justification/explanation

This principle deals with the drawing of well-justified conclusions on the basis of all the information collected during an experiment. In an experimental procedure, it is not sufficient to collect as much precise data as possible, but it is necessary to look for an explanation of these data. Therefore, not only the drawn conclusions must be strongly supported, but also all the data from an experiment should be interpreted in order to derive the correct implications. Usually, in physics, experiments are considered "good" when they are grounded in existing theories and use an apparatus that measures the quantities of interest with sufficient accuracy and precision (Franklin 1981). In this case, experiments are *theory laden* in that the terms and the used apparatus are dependent on existing theory. One of the important roles of experiments in physics is to test theories (Franklin 2009): as already discussed, an experiment can confirm an existing theory or can show that a new theory is required, either by showing that an accepted theory is incorrect or by exhibiting a new phenomenon that calls for an explanation. Therefore, the well-founded and comprehensive theories of physics provide the conceptual framework in which experiments are designed and realized. At the same time, theories are tested by experiments that, in some cases, can show that they are inadequate or incomplete. The difficult part concerns the fact that experiments may not always give clear-cut

results: the so called *crucial experiments*, those that quickly decide in a definitive way between two or more competing theories, are exceptions. In all other cases, experiments are much more difficult to interpret and, as a consequence, complex to explain.

Physicists, however, have a reasonable belief in experiments and in their results: how is that possible? This question has been addressed by many philosophers of science. Among those, Hacking (1983) proposed an elaborated answer that we believe can be valuable also for mobile robotics. Hacking sets out a number of strategies for believing in experiments. For example, he stresses the crucial role of *intervention* in the experimental practice. He considers in particular the case of microscopes used in experiments (Hacking 1981). For instance, in looking at a cell through a microscope, one may inject some fluid into the cell, in order to change the cell color when the operation is done. If one actually observes the predicted effect of this intervention, his/her beliefs in the proper operation of the experimental apparatus and in the data collected with it are strengthened. Of course this cannot be the only strategy. *Independent confirmation*, namely that the same data can be collected by different microscopes, is another strategy that may be combined with the first one. But what happens when an experiment can be performed with only one type of apparatus or when the intervention is very difficult, if not impossible? In such cases other strategies for believing experiments are possible. The experimental apparatus can reproduce known phenomena such that, in case of success, the belief in its proper working is enhanced. Other strategies include: the reproduction of artifacts that are known in advance, the elimination of alternative explanations of the results, the use of the results themselves to argue for their validity, the use of an independently well-corroborated theory of the phenomenon to explain the results, the use of an apparatus based on a well-corroborated theory, the adoption of statistical arguments. Of course it is very unlikely that these strategies can be adopted all together; usually just some of them are at disposal. They provide good reasons for believing in experimental results, even if they do not guarantee that the results are always correct. In the history of science, there are many cases in which these strategies have been adopted, but experimental results eventually turned out to be incorrect. Experiments are fallible: this does not mean that it is not possible to reach well-justified conclusions, but that these conclusions are not guaranteed once and for all. Among the above strategies, statistical methods have been already considered in computer engineering, in particular in algorithmics. For example, Barr et al. (1995) adopt statistical Design Of Experiments (DOE) to ensure that collected data can be analyzed by statistical methods to reach valid conclusions. DOE is based on the principles of replication (repeating tests), randomization (performing tests in random

order to offset non-included factors), and blocking (eliminating the influence of known, but extraneous, factors). It is worth remembering, however, that also in this more pragmatic engineering field the correct application of these statistical techniques cannot guarantee the correctness of experimental results.

4 Principles for experiments in localization and mapping

The analysis of the previous section has highlighted three important principles on which experimental activities in science and engineering are based. A question that arises is: to what extent are the experimental activities in mobile robot localization and mapping fulfilling these principles? In the following, we attempt to provide an answer. Before that, we remark that the three principles, and the related discussions, are closely interconnected and that their division here is purely for presentation purposes. Table 2 summarizes the main relations between aspects of Sect. 2 and principles of Sect. 3.3.

4.1 The comparison principle

Some of the aspects that have been identified in Sect. 2 as characterizing the current experimental activities of robotic localization and mapping appear suited to contribute to the

comparison principle. For example, the increasing use of publicly available data sets (like the Victoria Park data set, Guivant et al. 2000, and, more generally, those provided by RADISH, Howard and Roy 2003) sets a common ground for comparing the performance of different systems. With the imminent conclusion of the European project Rawseeds (2006), the availability of data sets for localization and mapping is expected to further increase.

Also the quantities that can be considered for comparison are somehow emerging from the current practice. These quantities include (see Sect. 2.3): computational complexity, computational time, memory usage, precision, and accuracy. The first three quantities can be used to describe the intrinsic performance of the systems, namely their efficiency. The last two quantities can be used to describe the extrinsic performance of the systems, namely their effectiveness in solving the problems of localization and mapping. However, computational time should not be considered as a decisive element of comparison, especially when the difference in performance is less than one second (Johnson 2002). This is because the systems could have been run on computers with slightly different configurations (e.g., with different system load) and, more importantly, because the technological evolution makes small differences in times vanishing in very few years.

A good news is that, from the picture sketched in Sect. 2, the interest of researchers in comparing their proposed systems with alternative ones is increasing. However, it is a fact

Table 2 Main relations between trends in experimental activities in mobile robot localization and mapping (Sect. 2) and principles of an experimental methodology (Sect. 3.3)

| | Principles of an experimental methodology | | |
|--|---|-----------------------------------|---------------------------|
| | Comparison | Reproducibility and repeatability | Justification/explanation |
| Purpose of experiments | | | |
| Demonstrating that system works | | | X |
| Getting insights on the behavior of the system and assessing limits of applicability | | X | X |
| Comparing the system with competing ones | X | | |
| Data sets | | | |
| Publicly available instances and code | X | X | |
| Use of different environments | | X | X |
| Measured quantities | | | |
| Computational complexity | X | | X |
| Computational time/Memory usage | X | X | |
| Profiling of the total time | X | X | X |
| Precision | X | X | X |
| Accuracy | X | X | X |
| Robustness | X | X | X |
| Report anomalies in performance | | X | |

that this comparison is not easy. As remarked by Johnson (2002), there are three ways to compare performance of algorithms, in decreasing order of appeal:

1. use the same code that was used in the previous experiments,
2. develop a comparable implementation, starting from the description provided in papers and reports,
3. compare the results with those obtained in other papers.

Although it is not always clear from the papers how data from other systems have been obtained, most comparative experiments in localization and mapping appear to adopt the second way (Frese et al. 2005; Minguez et al. 2006). Incentives to make code publicly available can push the comparison between different systems toward the first way. A remarkable initiative in this direction is OpenSLAM (Stachniss et al. 2007).

Finally, when comparing different systems, it is often not enough to show that one is performing better than the other ones, but this conclusion can be made stronger by explaining why this happens (Johnson 2002). This is sometimes attempted by resorting to the profiling of total time (see Sect. 2.3).

4.2 The reproducibility and repeatability principle

Reproducibility and repeatability constitute the second experimental principle that researchers in mobile robotics should consider in designing experimental activities. Due to their tight connection, and despite their slightly different desiderata, most of the experimental activities we surveyed in Sect. 2 support both principles.

Current practice shows an increasing awareness of the importance of reproducibility and repeatability as demonstrated by the detailed description of experimental procedures reported in recent papers (Grisetti et al. 2007; Minguez et al. 2006). These papers are not limited to the description of sketches of the proposed algorithms and to a qualitative evaluation with respect to inaccessible data sets. Instead, authors often distribute code and/or problem instances (Grisetti et al. 2007; Montemerlo et al. 2003), allowing the repeatability of the experiment, or, with an even more mature approach, there are tentatives to understand which parameters influence the system (Grisetti et al. 2007; Frese et al. 2005), allowing the implementation of similar experiments that should draw the same conclusions (reproducibility). This tendency requires the design of a set of so called *exploratory experiments* with the goal of understanding which implementation details, parameter settings, heuristics, and data structures affect the quality of the solution and should, therefore, be reported to the readers (Johnson 2002). Reproducing experiments is possible only if authors highlight to other researchers which are the sensitive parameters that must be carefully controlled and tuned.

Some papers have started to provide results in this direction, such as those by Grisetti et al. (2007) and Frese et al. (2005). These recent efforts are of great importance as they allow to achieve reproducibility of experimental results. Repeatability is, on the other hand, more difficult to obtain as experiments with real robots are both time consuming and subject to several hardware failures.

Another important issue in reproducibility is defining which are the quantities used for evaluation and, therefore, to state when a new experiment has been able to replicate the original results. These quantities are emerging from the current practice and include running time, either total (Montemerlo et al. 2003) or per single operation (Davison 2003; Davison et al. 2007; Grisetti et al. 2007), and number of iterations, a more reproducible quantity that is machine-independent (Frese et al. 2005; Minguez et al. 2006).

A solid experimentation should be based on results involving several data sets, referring to different kinds of environments (like indoor offices, indoor open spaces, outdoor crowded streets, outdoor parking lots, ...). In this way, it is possible to avoid drawing wrong conclusions or conclusions that are hardly reproducible. Reaching such a confidence in conclusions is a preliminary achievement for the repeatability of the experiment that, as pointed out in Sect. 3.3, means to achieve the same conclusions with similar, but not identical, settings. While most of the papers still deal with a single environment, often university buildings, there are some exceptions that use more data sets (Frese et al. 2005; Grisetti et al. 2007; Liu and Thrun 2003; Paz et al. 2008).

Several papers also deal with result variability of randomized algorithms, executing multiple runs of the same instance and reporting average values and standard deviations. Reporting only the best results does not support reproducibility, as it amounts to sample from the tail of the distribution, making results less likely to be replicated than the average (Johnson 2002). Example of good approaches to variability are in Minguez et al. (2006), where authors show the fraction of cases within a certain level of accuracy, and the χ^2 -test analysis presented by Frese et al. (2005).

A last important issue that has been highlighted in Sect. 3.3 is that science, as experimental activity, cannot be carried out in isolation. With the creation of repositories of freely available code, like OpenSLAM (Stachniss et al. 2007), and data sets, like RADISH (Howard and Roy 2003), the SLAM community has recently demonstrated awareness of the importance of other researchers' comments and critiques. Several papers (Grisetti et al. 2007; Montemerlo et al. 2003) are also proposing and using standard data sets, such as Victoria Park (Guivant et al. 2000), that are becoming *de facto* benchmarks. Even if not part of a formal design, results on standard benchmarks provide valuable points of reference, even in non-comparative settings but only for purposes of reproducibility. Another

aspect of sharing useful knowledge is reporting anomalies in performance. For example, Grisetti et al. (2007) discuss some critical situations, while some limits of applicability are presented in Neira and Tardos (2001). Reporting anomalies does not only allow to highlight which issues deserve further study in the future, but it is also a key for reproducibility, as it defines the boundaries of the domain where the drawn conclusions hold.

4.3 The justification/explanation principle

The third principle that is fundamental for experiments is the justification/explanation principle. Almost all the papers we have considered in Sect. 2 present aspects that address this principle. As for the previous principles, data sets are a decisive aspect for deriving well-justified conclusions about the behavior of a system. The system should have demonstrated to perform correctly over a significant number of different data sets and not just in a single instance in the authors' lab. Without any effort of dealing with more data sets, it is often easy to fall in the error of tuning the parameters on the specific instance so that there is no confidence in stating that the conclusions are true and not an artifact of the experimental setup. That said, the number of papers using several data sets is increasing (Frese et al. 2005; Grisetti et al. 2007; Liu and Thrun 2003; Paz et al. 2008).

Another aspect that appears important here is the robustness of the results. As we have seen in Sect. 2, this aspect is related to the correct behavior of the system when the magnitude of the errors (affecting the position of the robots, the data they perceive, ...) increase. It is evident that the more robust the system, the stronger the conclusions on its behavior. The correct behavior of a localization or mapping system is usually verified according to the ground truth, when available, or according to visual inspections of the produced maps. In the first case, some numerical values, like accuracy, can be used to support and strengthen the conclusions. In the second case, the conclusions are based on weaker reasons, which basically come down to the experience of the human evaluator, to his/her knowledge of the mapped environment, and to his/her ability to compare topologically the map and the real environment. Here a paradox comes into scene: the availability of ground truth means that the localization or mapping problem is already solved. Put it in another way, to show that a system works we compare its results with the ground truth on instances for which it is available, but what we really need is that the system works on instances for which ground truth is not available. Note that this aspect is related to the difficulty to determine a significant sample of environments in which robots operate and to generalize the results obtained in an environment to other environments. Strategies proposed by Hacking and discussed in Sect. 2.3 may play a role in further reasoning on this problem.

5 Conclusions

In this paper, we have taken an interdisciplinary perspective to compare experimental activities performed in science with experimental activities performed in mobile robot localization and mapping. We have shown that, since the aim of science, and of physics in particular, is to study natural reality and the aim of robotics is to design artificial systems, the role played by experiments is accordingly different. However, there are some principles that are underlying all experimental activities and that should be considered also in mobile robotics. Our aim is that the discussion we have presented can contribute to the definition of a stable, agreed-upon, and effective experimental methodology for localization and mapping. We do not attempt to make any definite proposal, but we just list some issues that we consider relevant for this definition.

- From the analysis of Sect. 4, it emerges that some localization and mapping works are addressing in a more and more convincing way the general principles characterizing experimental activities (comparison, reproducibility and repeatability, and justification/explanation principles). A current limitation is that, although the *union* of experimental activities conducted in the sample papers we surveyed well covers the three principles, their *intersection* is almost a null set. Put it in another way, the problem of experimental activity in mobile robotics is complex and involves multiple dimensions, from which each researcher selects a subset of dimensions, according to his/her current needs. This can be due to the relatively young discussion on good experimental practices that has not yet brought to a stable experimental methodology. However, current trends seem to go in the right direction.
- A mobile robot is a complex system. In most part of the above discussion, we implicitly supposed that single components, like localization and mapping components, can be considered as independent from the rest of the system. The relationships between the single components and the whole system and their impact on experiments require further consideration.
- Proposing an experimental methodology for mobile robot localization and mapping involves also rethinking how papers are written, what are the requirements for publishing in conferences and journals, how project proposals are prepared and selected, and so on. Just as an example, we deem that papers presenting negative results should not be banned, as it seems to happen if one looks at the literature of the latest years. Instead, negative results can contribute to the understanding of systems, according to the justification/explanation principle and to what happens in other disciplines.

– While experiments in physics are used to confirm/refute a theory or to discover a new theory, in robotics it is not clear what plays the role of theory. Interesting speculations could arise from reflecting on this issue.

In conclusion, we think that, in defining good experimental practices for mobile robot localization and mapping, it is useful to have a look at the long tradition of studies on experimental practices in science and engineering. With this paper we have just started working in this direction, by surveying some current trends in mobile robotics and comparing them with well-established principles of experiments in science and engineering.

Acknowledgements The authors gratefully thank Alberto Pretto and Dario Lodi Rizzini for helping in selecting the sample papers surveyed in Sect. 2 and Nicola Basilico and Stefano Ghidoni for providing useful suggestions.

References

- Amigoni, F. (2008). Experimental evaluation of some exploration strategies for mobile robots. In *IEEE int. conf. on robotics and automation* (pp. 2818–2823).
- Amigoni, F., Gasparini, S., & Gini, M. (2007). Good experimental methodologies for robotic mapping: A proposal. In *IEEE int. conf. on robotics and automation* (pp. 4176–4181).
- Barr, R. S., Golden, B. L., Kelly, J. P., Resende, M. G. C., & Stewart, W. R. (1995). Designing and reporting on computational experiments with heuristic methods. *Journal of Heuristics*, 1, 9–32.
- Bonsignorio, F., Hallam, J., & del Pobil, A.P. (2007). Good experimental methodology—GEM guidelines. <http://www.heeronrobots.com/EuronGEMSig/Downloads/GemSigGuidelinesBeta.pdf>.
- Comport, A. I., Malis, E., & Rives, P. (2007). Accurate quadrifocal tracking for robust 3d visual odometry. In *IEEE int. conf. on robotics and automation* (pp. 40–45).
- Cromer, A. (1993). *Uncommon sense*. New York: Oxford University Press.
- Davison, A. J. (2003). Real-time simultaneous localisation and mapping with a single camera. In *IEEE int. conf. on computer vision* (pp. 1403–1410).
- Davison, A. J., Reid, I. D., Molton, N. D., & Stasse, O. (2007). Monoslam: real-time single camera slam. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6), 1052–1067.
- EURON GEM Sig (2007). <http://www.heeronrobots.com/EuronGEMSig/>.
- Franklin, A. (1981). What makes a ‘good’ experiment? *British Journal for the Philosophy of Science*, 32(4), 367–379.
- Franklin, A. (2009). Experiment in physics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. <http://plato.stanford.edu/archives/spr2009/entries/physics-experiment/>.
- Frese, U., Larsson, P., & Duckett, T. (2005). A multilevel relaxation algorithm for simultaneous localization and mapping. *IEEE Transactions on Robotics*, 21(2), 196–207.
- Grisetti, G., Stachniss, C., & Burgard, W. (2007). Improved techniques for grid mapping with rao-blackwellized particle filters. *IEEE Transactions on Robotics*, 23(1), 34–46.
- Grosser, M. (1962). *The discovery of Neptune*. Cambridge: Harvard University Press.
- Guivant, J., Nebot, E., & Baiker, S. (2000). Autonomous navigation and map building using laser range sensors in outdoor applications. *Journal of Robotic Systems*, 17(10), 565–583.
- Gutmann, J. S., & Konolige, K. (1999). Incremental mapping of large cyclic environments. In *IEEE int. symp. on computational intelligence in robotics and automation* (pp. 318–325).
- Hacking, I. (1981). Do we see through a microscope? *Pacific Philosophical Quarterly*, 62, 305–322.
- Hacking, I. (1983). *Representing and intervening. Introductory topics in the philosophy of natural science*. Cambridge: Cambridge University Press.
- Hahnel, D., Burgard, W., Fox, D., & Thrun, S. (2003). An efficient fastslam algorithm for generating maps of large-scale cyclic environments from raw laser range measurements. In *IEEE/RSJ int. conf. on intelligent robots and systems* (pp. 206–211).
- Hanks, S., Pollack, M. E., & Cohen, P. R. (1993). Benchmarks, test beds, controlled experimentation, and the design of agent architectures. *AI Magazine*, 14(4), 17–42.
- Howard, A., & Roy, N. (2003). The robotics data set repository (radish). <http://radish.sourceforge.net/>.
- Johnson, D. (2002). A theoretician’s guide to the experimental analysis of algorithms. In M. H. Goldwasser, D. S. Johnson, & C. C. McGeoch (Eds.), *Data structures, near neighbor searches, and methodology: fifth and sixth dimacs implementation challenges* (pp. 215–250). Providence: American Mathematical Society.
- Leonard, J. J., & Durrant-Whyte, H. F. (1991). Simultaneous map building and localization for an autonomous mobile robot. In *IEEE/RSJ int. conf. on intelligent robots and systems* (pp. 1442–1447).
- Liu, Y., & Thrun, S. (2003). Results for outdoor-slam using sparse extended information filters. In *IEEE int. conf. on robotics and automation* (pp. 1227–1233).
- Lu, F., & Milios, E. (1997). Globally consistent range scan alignment for environment mapping. *Autonomous Robots*, 4(4), 333–349.
- Minguez, J., Montesano, L., & Lamiroux, F. (2006). Metric-based iterative closest point scan matching for sensor displacement estimation. *IEEE Transactions on Robotics*, 22(5), 1047–1054.
- Montemerlo, M., Thrun, S., Koller, D., & Wegbreit, B. (2003). Fastslam 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges. In *Int. joint conf. on artificial intelligence* (pp. 1151–1156).
- Montgomery, D. (2005). *Design and analysis of experiments*. New York: Wiley.
- Neira, J., & Tardos, J. D. (2001). Data association in stochastic mapping using the joint compatibility test. *IEEE Transactions on Robotics and Automation*, 17(6), 890–897.
- Newman, P., Leonard, J., Tardos, J. D., & Neira, J. (2002). Explore and return: experimental validation of real-time concurrent mapping and localization. In *IEEE int. conf. on robotics and automation* (pp. 1802–1809).
- Newman, P., Cole, D., & Ho, K. (2006). Outdoor slam using visual appearance and laser ranging. In *IEEE int. conf. on robotics and automation* (pp. 1180–1187).
- Paz, L. M., Pinies, P., Tardos, J. D., & Neira, J. (2008). Large-scale 6-dof slam with stereo-in-hand. *IEEE Transactions on Robotics*, 24(5), 946–957.
- Permis (2000). Performance metrics for intelligent systems. http://www.isd.mel.nist.gov/research_areas/research_engineering/Performance_Metrics/index.htm.
- Rawseeds (2006). <http://rawseeds.elet.polimi.it/>.
- RoSta (2007). Robot standards and reference architectures. <http://www.robot-standards.eu/>.
- Stachniss, C., Frese, U., & Grisetti, G. (2007). Openslam.org. <http://www.openslam.org/>.
- Thrun, S., Burgard, W., & Fox, D. (2005). *Probabilistic robotics*. Cambridge: The MIT Press.
- Westfall, R. (1971). *The construction of modern science. Mechanisms and mechanics*. New York: Wiley.



Francesco Amigoni was born in Soresina (CR), Italy, on December 2, 1971. He got the Laurea degree in Computer Engineering from the Politecnico di Milano in 1996 and the PhD degree in Computer Engineering and Automatica from the Politecnico di Milano in 2000. From December 1999 to September 2000 he has been a visiting scholar at the Computer Science Department of the Stanford University (USA). From February 2002 to April 2007 he has been an assistant professor at the Dipartimento di Elettronica e In-

formazione of the Politecnico di Milano. From May 2007 he is an associate professor at the Dipartimento di Elettronica e Informazione of the Politecnico di Milano. His main research interests include: agents and multiagent systems, mobile robotics, and the philosophical aspects of artificial intelligence. He is author of about 60 papers published on international journals, books, and conference proceedings.



Monica Reggiani graduated from the University of Parma in 1997 and obtained the PhD in Computer Engineering in 2001 from the same University. In 1999–2000 she was a visiting scholar at the Los Alamos National Laboratory working on parallel and distributed computing. From 2001 to 2005 she was a research associate at the Dipartimento di Ingegneria dell'Informazione, University of Parma. In 2006–2007 she was a research associate in the European project "Xpero: learning by experimentation" with the Uni-

versity of Verona. In 2007 she has been a visiting professor at the University of Applied Sciences Bonn-Rhein-Sieg. Currently, she is an assistant professor at the Dipartimento di Tecnica e Gestione dei Sistemi Industriali, University of Padua. Her areas of interest include cognitive robotics, motion planning, programming by demonstration, and autonomous robots.



Viola Schiaffonati got the Laurea degree in Philosophy from Università degli Studi di Milano in 1999 and the PhD in Philosophy of Science from Università di Genova in 2004. She has been visiting scholar at the Department of Philosophy of the University of California at Berkeley during the academic year 2000/01 and visiting researcher at the Suppes Center for the Interdisciplinary Study of Science and Technology of the Stanford University in October 2005. Currently she is a temporary researcher at the Diparti-

mento di Elettronica e Informazione of the Politecnico di Milano and contract professor in the Facoltà di Ingegneria dei Sistemi and in the Facoltà di Ingegneria dell'Informazione of the Politecnico di Milano. Her main research interests include: philosophical foundations of artificial intelligence, formal approaches to the philosophy of science, and epistemological issues of computational science.