2

Introduction to Scientific Methods in Mobile Robotics

Summary. This chapter introduces the main topic of this book, identifies the aims and objectives and describes the background the material presented in this book.

2.1 Introduction

The behaviour of a mobile robot emerges from the relationship and interaction between the robot's control code, the environment the robot is operating in, and the physical makeup of the robot. Change any of these components, and the behaviour of the robot will change.

This book is concerned with how to characterise and model, "identify", the behaviour emerging from the interaction of these three components. Is the robot's behaviour predictable, can it be modelled, is it stable? Is this behaviour different from that one, or is there no significant difference between them? Which programs performs better (where "better" is some measurable criterion)?

To answer these questions, we use methods taken from dynamical systems theory, statistics, and system identification. These methods investigate the dynamics of robot-environment interaction, and while this interaction *is* also governed by the control program being executed by the robot, they are not suited to analyse *all* aspects of robot behaviour. For example, dynamical systems theory will probably not characterise the relevant aspects of the behaviour of a robot that uses computer vision and internal models to steer towards one particular location in the world. In other words, the methods presented in this book are primarily concerned with dynamics, not with cognitive aspects of robot behaviour.

This book aims to extend the way we conduct autonomous mobile robotics research, to add a further dimension: from a discipline that largely uses iterative refinement and trial-and-error methods to one that is based on testable hypotheses, that makes predictions about robot behaviour based on a theory of robot-environment interaction. The book investigates the mechanisms that give rise to robot behaviour we observe: why does a robot succeed in certain environments and fail in others? Can we make accurate predictions as to what the robot is going to do? Can we *measure* robot behaviour?

Although primarily concerned with physical mobile robots, operating in the real world, the mechanisms discussed in this book can be applied to all kinds of "behaving agents", be it software agents, or animals. The underlying questions in all cases are the same: can the behaviour of the agent be *measured quantitatively*, can it be modelled, and can it be predicted?

2.1.1 A Lecture Plan

This book is the result of undergraduate and postgraduate courses in "Scientific Methods in Mobile Robotics" taught at the University of Essex, the Memorial University of Newfoundland, the University of Palermo and the University of Santiago de Compostela. The objective of these courses was to introduce students to fundamental concepts in scientific research, to build up knowledge of the relevant concepts in philosophy of science, experimental design and procedure, robotics and scientific analysis, and to apply these specifically to the area of autonomous mobile robotics research. Perhaps it is easiest to highlight the topics covered in this book through this sequence of lectures, which has worked well in practice:

1. Introduction (Chapter 2):
   - Why is scientific method relevant to robotics? How can it be applied to autonomous mobile robotics?
   - The robot as an analog computer (Section 2.3)
   - A theory of robot-environment interaction (Section 2.4)
   - The role of quantitative descriptions (Section 2.4.2)
   - Robot engineering vs robot science (Section 2.5)
2. Scientific Method (Section 2.6):
   - Forming hypotheses (Section 2.6.2)
   - Experimental design (Section 2.6.3)
   - Traps, pitfalls and countermeasures (Section 2.6.3)
3. Introduction to statistical descriptions of robot-environment interaction:
   - Normal distribution (Sections 3.2 and 3.3.2)
4. Parametric tests to compare distributions:
   - T-test (Sections 3.3.4 and 3.3.5)
   - ANOVA (Section 3.3.6)
5. Non-parametric tests I:
   - Median and confidence interval (Section 3.4.1)
   - Mann-Whitney U-test (Section 3.4.2)
6. Non-parametric tests II:
   - Wilcoxon test for paired observations (Section 3.4.3)
   - Kruskal-Wallis test (Section 3.4.4)
   - Testing for randomness (Section 3.5)

7. Tests for a trend:
   - Linear regression (Section 3.6.1)
   - Pearson's r (Section 3.6.2)
   - Spearman rank correlation (Section 3.7.1)
8. Analysing categorical data (Section 3.8):
   - χ^2 analysis (Section 3.8.1)
   - Cramer's V (Section 3.8.2)
   - Entropy based methods (Section 3.8.3)
9. Dynamical systems theory and chaos theory (Chapter 4):
   - Phase space (Section 4.2.1)
   - Degrees of freedom of a mobile robot (Section 4.2.1)
   - The use of quantitative descriptions of phase space in robotics (Section 2.4.2)
   - Reconstruction of phase space through time-lag embedding (Section 4.2.3)
10. Describing robot behaviour quantitatively through phase space analysis (Section 4.3)
11. Quantitative descriptors of attractors:
    - Lyapunov exponent (Section 4.4)
    - Prediction horizon (Section 4.4.2)
    - Correlation dimension (Section 4.5)
12. Modelling of robot-environment interaction (Chapter 6)
13. ARMAX modelling (Section 6.4.3)
14. NARMAX modelling (Section 6.5):
    - Environment identification (Section 6.6)
    - Task identification (Section 6.7)
    - Sensor identification (Section 6.8)
15. Comparison of behaviours (Section 6.9)
16. Summary and conclusion (Chapter 7)

2.2 Motivation: Analytical Robotics

The aim of this book is to throw some light light on the question "what happens when a mobile robot — or in fact any agent — interacts with its environment?". Can predictions be made about this interaction? If models *can* be built, can they be used to design autonomous mobile robots off-line, like we are now able to design buildings, electronic circuits or chemical compounds without applying trial-and-error methods? Can models be built, and can they be used to hypothesise about the nature of the interaction? Is the process of robot-environment interaction stochastic or deterministic?

Why are such questions relevant? Modern mobile robotics, using autonomous mobile robot with their own on-board power supply, sensors and computing equipment, is a relatively new discipline. While as early as 1918 a light-seeking robot was built by John Hays Hammond [Loeb, 1918, chapter 6], and W. Grey Walter built mobile robots that *learnt* to move towards a light source by way of instrumental conditioning in the 1950s [Walter, 1950, Walter, 1951], "mass" mobile robotics really only began in the 1980s. As in all new disciplines, the focus was initially on the engineering aspects of getting a robot to work: which sensors can be used in a particular task, how do they need to be preprocessed and interpreted, which control mechanism should be used, *etc*. The experimental scenario used was often one of iterative refinement: a good first guess at a feasible control strategy was implemented, then tested in the target environment. If the robot got stuck, failed at the task *etc.*, the control code would be refined, then the process would be repeated until the specified task was successfully completed in the target environment.

A solution obtained in this manner constituted an "existence proof" — it was proven that a particular robot could achieve a particular task under a particular set of environmental conditions. These existence proofs were good achievements, because they demonstrated clearly that a particular behaviour or competence could be achieved, but they lacked one important property: generality. That a robot could successfully complete a navigational route in one environment did not imply that it could do it anywhere else. Furthermore, the experimenter did not really know *why* the robot succeeded. Success or failure could not be determined to a high degree of certainty *before* an experiment. Unlike building bridges, for instance, where civil engineers are able to predict the bridge's behaviour before it is even built, roboticists are unable to predict a robot's behaviour before it is tested.

Perhaps the time has come for us to be able to make some more general, theoretical statements about what happens in robot-environment interaction. We have sophisticated tools such as computer models (see Chapter 6) and analysis methods (see Chapter 4), which can be used to develop a *theory* of robot-environment interaction. If this research wasn't so practical, involving physical mobile robots doing something in the real world, I would call the discipline "theoretical robotics". Instead, I use the term "analytical robotics".

In addition there are benefits to be had from a theory of robot-environment interaction: the more theoretical knowledge we have about robot-environment interaction, the more accurate, reliable and cheap will the robot and controller design process be. The more we know about robot-environment interaction, the more focused and precise will our hypotheses and predictions be about the outcome of experiments. This, in turn, will increase our ability to detect rogue experimental results and to improve our experimental design. Finally, the better understood the process of robot-environment interaction, the better we are able to report experimental results, which in turn supports independent replication and verification of results: robotics would advance from an experimental discipline to one that embraces scientific method.

The aim of this book, therefore, is to understand robot-environment interaction more clearly, and to present abstracted, generalised representations of that interaction — a theory of robot-environment interaction.

2.3 Robot-Environment Interaction as Computation

The behaviour of a mobile robot cannot be discussed in isolation: it is the result of properties of the robot itself (physical aspects — the "embodiment"), the environment ("situatedness"), and the control program (the "task") the robot is executing (see Figure 2.1). This triangle of robot, task and environment constitutes a complex, interacting system, whose analysis is the purpose of any theory of robot-environment interaction.

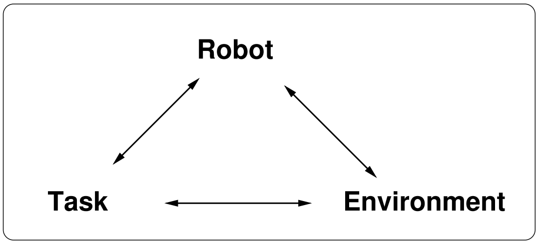

Figure 2.1. The fundamental triangle of robot-environment interaction

Rather than speaking solely of a robot's behaviour, it is therefore necessary to speak of *robot-environment interaction*, and the robot's behaviour resulting thereof.

A mobile robot, interacting with its environment, can be viewed as performing "computation", "computing" *behaviour* (the output) from the three inputs *robot morphology*, *environmental characteristics* and *executed task* (see Figure 2.2).

Similar to a cylindrical lens, which can be used to perform an analog computation, highlighting vertical edges and suppressing horizontal ones, or a camera lens computing a Fourier transform by analog means, a robot's behaviour — for the purposes of this book, and as a first approximation, the mobile robot's trajectory — can be seen as emergent from the three components shown in Figure 2.1: the robot "computes" its behaviour from its own makeup, the world's makeup, and taking into account the program it is currently running (the task).

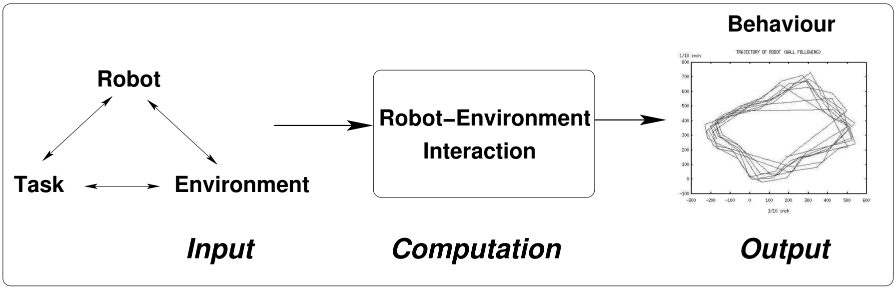

Figure 2.2. Robot-environment interaction as computation: Behaviour (the output) is computed from the three inputs robot morphology, task and environmental properties

2.4 A Theory of Robot-Environment Interaction

2.4.1 Definition

When referring to "theory", we mean a coherent body of hypothetical, conceptual and pragmatic generalisations and principles that form the general frame of reference within which mobile robotics research is conducted.

There are two key elements that make a theory of robot-environment interaction useful, and therefore desirable for research:

1. A theory will allow the formulation of hypotheses for testing. This is an essential component in the conduct of "normal science" [Kuhn, 1964].
2. A theory will make predictions (for instance regarding the outcome of experiments), and thus serve as a safeguard against unfounded or weakly supported assumptions.

A theory retains, in abstraction and generalisation, the essence of what it is that the triple of robot-task-environment does. This generalisation is essential; it highlights the important aspects of robot-environment interaction, while suppressing unimportant ones. Finally, the validity of a theory (or otherwise) can then be established by evaluating the predictions made applying the theory.

Having theoretical understanding of a scientific discipline has many advantages. The main ones are that a theory allows the generation of hypotheses and making testable predictions, but there are practical advantages, too, particularly for a discipline that involves the design of technical artefacts. For instance, theory supports off-line design, *i.e.* the design of technical artefacts through the use of computer models, simulations and theory-based calculations.

2.4.2 The Role of Quantitative Descriptions of Robot-Environment Interaction

Measurement is the backbone of science, and supports:

- The precise documentation of experimental setups and experimental results
- The principled modification of experimental parameters
- Independent verification of experimental results
- Theoretical design of artefacts without experimental development
- Predictions about the behaviour of the system under investigation

We have argued that robot behaviour emerges from the interaction between robot, task and environment. Suppose we were able to measure this behaviour quantitatively. Then, if any two of the three components shown in Figure 2.1 remain unaltered, the quantitative performance measure will characterise the third, modified component. This would allow the investigation of, for instance:

- The effect of modifications of the robot
- The influence of the robot control program on robot behaviour
- The effect of modifications to the environment on the overall behaviour of the robot

This is illustrated in Figure 2.3: the quantitative measure of the robot's behaviour (the dependent variable) changes as some experimental parameter (the independent variable) changes, and can therefore be used to describe the independent variable. For the point γ in Figure 2.3, for example, the quantitative performance measure has a global maximum.

Chapter 4 in particular addresses the question of how robot-environment interaction can be characterised quantitatively, and how such quantitative measures can be used to determine the influence of i) a change in the robot controller, and ii) a change of environment.

Current mobile robotics research practice not only differs from that of established disciplines in its lack of theories supporting design, but also in a second aspect: independent replication and verification of experimental results in mobile robotics is, as yet, uncommon. While in sciences such as biology or physics, for instance, reported results are only taken seriously once they have been verified independently a number of times, in robotics this is not the case. Instead, papers often describe experimental results obtained in specific environment, under specific experimental conditions. These experiments therefore are "existence proofs" — the demonstration that a particular result can be achieved — but they do not state in general terms under which conditions a particular result can be obtained, nor which principles underlie the result. Existence proofs are useful, they demonstrate that something can be achieved, which is an important aspect of science, but they do not offer general principles and theories.

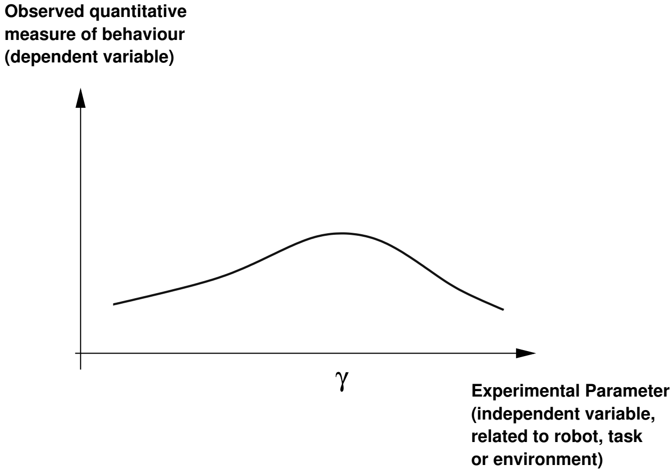

Figure 2.3. Illustration of a conceivable relationship between quantitative performance measure and experimental parameter

We argue that mobile robotics research is now at a stage where we should move on from existence proofs to a research culture that habitually includes independent replication and verification of experiments.

Theories, experimental replication and experimental verification all depend crucially on *quantitative* descriptions: quantitative descriptions are an essential element of the language of science. For these reasons this book presents several ways of describing robot-environment interaction quantitatively[1].

2.5 Robot Engineering *vs* Robot Science

Arguably, there are (at least) two independent objectives of robotics research: on the one hand, to create artefacts that are capable of carrying out useful tasks in the real world — for example industrial, service, transportation or medical robots, to name but a few, and on the other hand to obtain a theoretical understanding of the design issues involved in making those artefacts — for example sensor and actuator modelling, system identification (modelling of entire systems), or sensor, actuator and behaviour analysis. The former can be referred to as "robot engineering", the latter as "robot science". It is robot science that this book is mainly concerned with.

While robot engineering ultimately produces the "useful" artefacts, there is a lot that robot science can contribute to this process. Without theoretical understanding, any design process is largely dependent upon trial-and-error exper-

[1] A very informative article on quantitative measures of robot-environment interaction can be found in [Smithers, 1995].

imentation and iterative refinement. In order to design in a principled way, a hypothesis — a justified expectation — is needed to guide the design process. The hypothesis guides the investigation: results obtained are fed back into the process and brought into alignment with the theory, to lead to the next stage of the experimentation and design. The better the theory underlying the design process, the more effective and goal-oriented the design process will be. *Every* process of designing technical artefacts is based on some kind of assumptions (a "theory"), even if very little is known at all about the object being designed.

This is true for current mobile robotics research, too. When asked to design a wall-following robot, the designer will not start with an *arbitrary program*, but with a "reasonable guess", sensibly speculating on which sensors might be useful to achieve the desired behaviour, which general kind of control program will perform acceptably, *etc*. But, given our current understanding of robotics, he is unable to design the entire behaviour off-line!

Instead, mobile robotics researchers to-date are crucially dependent on trial-and-error procedures. A "reasonable prototype" has to be tested in the target environment, and refined based on observations and underlying theory ("hunch" is often the more appropriate term for such theories). Here is a practical example: to design the Roomba commercial robot floor cleaner (relying on very simple sensing, and not involving any sophisticated navigation), 30 prototypes had to be built over a period of 12 years [EXN, 2003]!

Theoretical understanding of robot-environment interaction, however, would address this issue, and support off-line design. But not only that: it would furthermore allow the analysis of an observed behaviour, and the refinement of existing mechanisms, based on established theoretical principles.

The argument this book makes, therefore, is this: a better theoretical understanding of the principles underlying a mobile robot's operation in its environment — a theory — will result in more effective, rigorous and goal-oriented development methods. These, in turn, will support robot engineering, leading to robots that are better able to achieve the tasks they are designed for.

2.6 Scientific Method and Autonomous Mobile Robotics

2.6.1 Introduction

Whether mobile robotics actually is a science or an engineering discipline, it undoubtedly benefits from clear, coherent and methodical research practice, and the following discussion should be relevant to both "science" and "engineering".

The discipline of mobile robotics is interested in developing artefacts (robots) that can carry out some useful task in a real world environment. However this is attempted, be it trial-and-error, methodical research or a mixture of both, the designer will rely on some previously acquired knowledge, perhaps inadvertently. This knowledge essentially constitutes a "theory". It is useful to analyse in more detail what the elements of this theory are, and how the theory can be improved — this is the purpose of this chapter.

2.6.2 Background: What is "Scientific Method"?

As stated earlier, the aim of this book is to open up new avenues of conducting research in mobile robotics, to move away from existence proofs and the need for iterative refinement, and to overcome the inability to design task-achieving robots off line. Before we look at some practical ways of applying scientific method to mobile robotics research, we'll look at a very broad summary of what has been understood by the term "scientific method" over the centuries. For a proper treatment of this topic, however, please see dedicated books on the subject (for example [Gillies, 1996, Harris, 1970, Gower, 1997].)

Sir Francis Bacon (1561 – 1626) first developed the theory of inductivism [Bacon, 1878], where the basic idea is this: first, a large number of observations regarding the subject under investigation is gathered. This includes "instances where a thing is present", "instances where a thing is not present", and "instances where a thing varies". The nature of the phenomenon under investigation is then determined by a process of eliminative induction. Almost mechanically, by gathering more and more information and ruling out impossible hypotheses, the truth is established. [Gillies, 1996] likens this inductive process to that of drawing a precise circle: impossible to achieve just using pen and paper, but very easy using the mechanical device of a compass. In a similar manner, scientific truths are to be discovered by the mechanical process of induction. The "problem of induction", however, is that the facts gathered can never be complete enough to fully justify the conclusions drawn, so that any hypotheses are in effect working hypotheses only, a first stab, so to speak.

The complete opposite view to Bacon's induction based on many observations is Karl Popper's argument that induction is a myth, because observation without theory is impossible [Popper, 1959, Popper, 1963, Popper, 1972]. In other words, there needs to be a theory first in order to observe, and obtaining a theory from a large body of observations alone is impossible. Simply "observing" cannot be done, the scientist needs to know what should be observed. This in turn requires the definition of a chosen task, a question, a problem — in other words: a hypothesis. Instead of inductivism, he proposed a theory of conjectures and refutations (falsificationism): the aim of scientific investigation is to refute a hypothesis, and all experimentation is geared towards that goal. If a hypothesis withstands all attempts of refutation, it is *tentatively* adopted as true, but not considered proven and true without doubt. The only truth that can be firmly established is that a theory is false, never that it is true.

How then does the scientific community accept or reject theories? Thomas Kuhn [Kuhn, 1964] differentiates between "normal science" and a situation of "scientific revolution". Normal science he describes as research firmly based on past scientific achievements or "paradigms". Paradigms here refer to theories that create avenues of enquiry, formulate questions, select methods and define relevant research areas — paradigms guide research. "Normal" scientific research aims to extend the knowledge within an existing paradigm, to match facts with theory, to articulate theory and to bring the existing theory into closer agreement with observed facts. It tends to suppress fundamental novelties that cannot be brought into agreement with existing paradigms. Normal science works within the accepted, existing paradigm, seeks to extend the knowledge the paradigm is revealing, and to "tie up loose ends" and plug gaps — Kuhn refers to this as "mopping up".

However, in the process of normal science increasingly discrepancies between fact and theory (anomalies) will be observed. There will be observations that cannot be explained at all with the existing theory, and there will be observations that appear to disagree with existing theory. These difficult cases tend to be ignored initially, but their weight and importance may increase until a point is reached at which the scientific community loses faith in the existing paradigm. A crisis has developed; it begins with a blurring of the existing paradigms, continues by the emergence of proposals for alternative paradigms, and eventually leads to a "scientific revolution", the transition form "normal" to extraordinary research. Eventually, the new paradigm is adopted by the majority of scientists and assumes the role of "normal" paradigm, and the process is repeated.

Scientific Research Methodology

As stated in the introduction, this book is no attempt to present an account of philosophy of science and its application to mobile robotics. When we refer to "scientific method", the emphasis is not on the philosophical foundations of research methodology.

Rather, it is on the procedure of conducting, evaluating and reporting research and its results; that is, the material practice of science, the "recipes". What is a good starting point for research? How do we design experiments, how do we document and assess the results? What do we adopt as a scientific research procedure within the community? These are the kinds of questions we should be able to answer before we conduct the actual research!

Forming Scientific Hypotheses

The starting point for any research is a hypothesis, a thesis. This hypothesis is a formally stated expectation about a behaviour that defines the purpose and the goals of a study; it therefore defines, explains and guides the research. Without a clear hypothesis in the beginning, it is virtually impossible to conduct good research, as it is virtually impossible to present results in a coherent and convincing way. The hypothesis, the question, is the foundation upon which the scientific argument is built. Obviously, an ambiguous question will result in an ambiguous answer, which is why the hypothesis is the most fundamental stage of scientific working.

To formulate the hypothesis clearly, it is useful to consider the following points (see also [Paul and Elder, 2004]):

1. What is the question addressed?
   - State it precisely
   - Can it be broken down into sub questions?
   - Is there one right answer to the question? Does it require reasoning from more than one point of view? Is it a matter of opinion?
2. What assumptions are you making?
   - Identify all assumptions clearly
   - Are they justifiable?
   - Do these assumptions affect the impartiality of your research?
   - Identify key concepts and ideas that shape the research. Are they reasonable?
3. Formulate a hypothesis
   - Is this hypothesis testable and falsifiable?
   - What outcome do you expect?
   - What would be the implications of the different possible outcomes of your experiment (*i.e.* is the question actually worth asking)?
   - Experimental design
4. Which experimental setup is suitable to investigate the question/hypothesis?
   - How is experimental data going to be collected?
   - How is experimental data going to be evaluated?
   - How much data is needed?

Hypotheses can be causal hypotheses, hypothesising about the causes of a behaviour, or descriptive, describing a behaviour in terms of its characteristics or the situation in which it occurs. Causal reasoning and causal models are very common in science, and guide experimental design, hypothesis formation and the formation of theories. Causal models guide scientific thinking so strongly that on occasions scientists even override the statistical information they receive, in favour of a causal model [Dunbar, 2003] (referred to as "confirmation bias" — "cold fusion" being a prominent example). In other words: the hypotheses guiding research can be so dominant that the scientist tries to generate results that confirms his initial hypothesis, rather than aiming to disprove a hypothesis (which is, according to Popper, what he should be doing!) [Klayman and Ha, 1987] — the tendency of trying to *confirm* a hypothesis, rather than refute it, is difficult to overcome. The temptation to conduct experiments that produce results predicted by the current hypothesis is very strong!

Popper argued that (due to the infinity of the universe) scientific hypotheses can *never* be verified (*i.e.* proven to be true) nor the probability of their veracity established, but that they can only be falsified, *i.e.* shown to be incorrect. He further argued that the most fundamental requirement for any scientific hypothesis must therefore be that the theory is open to tests and open to revision. In other words: it must be testable, and it must be falsifiable. If either of these conditions isn't met, the hypothesis will not support scientific investigation.

Popper was aware that it is possible to evade falsification by adopting "saving stratagems" (*e.g.* by modifying testability of a hypothesis), and therefore introduced the supreme rule that "the other rules of scientific procedure must be designed in such a way that they do not protect any statement in science from falsification" [Popper, 1959, p.54].

"The aim of science is to find satisfactory explanations, of whatever strikes us as being in need of explanation" [Popper, 1972, p. 191] — the hypothesis underlying the research ultimately defines the degree to which an explanation is satisfactory or not.

There are further criteria that distinguish "good" hypotheses from "bad" ones. Popper and Kuhn identify explanatory depth as a crucial aspect — which paradigm explains more phenomena? —, but increased verisimilitude is equally identified by Popper as an objective for forming hypotheses. In a survey article, summarising the views put forward by Kuhn, Lakatos and Laudan, [Nola and Sankey, 2000] state that "Scientists prefer a theory that

- Can solve some of the empirical difficulties confronting its rivals
- Can turn apparent counter-examples into solved problems
- Can solve problems it was not intended to solve
- Can solve problems not solved by its predecessors
- Can solve all problems solved by its predecessors, plus some new problems
- Can solve the largest number of important empirical problems while generating the fewest important anomalies and conceptual difficulties"

Hypotheses must be precise, rational (that is, possibly true and in agreement with what is already known) and parsimonious (that is, as simple as possible — but not simpler. William of Occam's razor — "entities are not to be multiplied beyond necessity" — is one expression of this principle). In summary, the hallmarks of a "good" scientific paradigm — which must be testable and falsifiable — are explanatory power, clarity and coherence.

How can scientific hypotheses be obtained? The most common sources are:

- Opinions, observations and experiences
- Existing research
- Theories
- Models

Karl Popper argued that scientific hypotheses are the product of brilliant creative thinking by the scientist (he refers to this as "creative intuition").

2.6.3 Experimental Design and Procedure

Experimental Design

Experimental design — the experimental procedure used, the observation mechanisms and the way results are interpreted — is the centre of any scientific investigation, and care is necessary when designing experiments. Is the chosen design suitable for investigating the hypothesis I am interested in? Is there a better way of achieving my objectives? Is the design feasible in practice, or does it offer insurmountable practical problems?

One of the most common types of scientific experiments aim to determine a relationship between two variables: one that is controlled by the experimenter (the independent variable, IV), and one that is dependent on it (the dependent variable, DV). The most common aim of experimentation is to establish how the DV changes in relation to the IV.

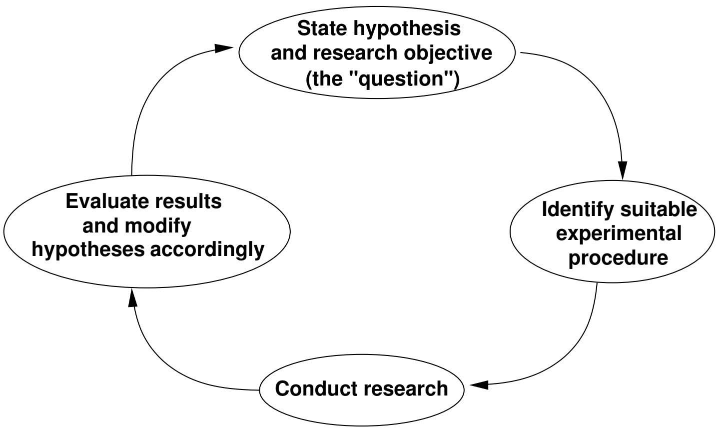

Figure 2.4. Experimental procedure

There are a number of aspects to be considered when designing an experiment (see also Figure 2.4):

- What is the question being asked? What is the hypothesis? *Every* scientific investigation is driven by the underlying question it has set out to answer. If this question is not formulated clearly, or even not formulated at all, the resulting research will be haphazard, ill focused without clear aim. Good research needs a clearly formulated objective!
- Sensitivity of the experiment. Is the experimental design sensitive enough to detect any causal relationship between DV and IV? Is it perhaps too sensitive, and will therefore amplify noise?

- Are there any confounding effects that introduce errors that hide any true effects (see below for a discussion of this point)?
- Which predictions can be made about the outcome of the experiment? Are there expectations, perhaps based on some prior understanding of the problem, that can be used to assess the eventual outcome of the experiment?
  Predictions are important, they serve as a sanity check, helping us to identify results that are highly unlikely, and to detect possible errors in the experimental design, procedure and evaluation.
- Observation. How is the experiment going to be observed, how are results going to be recorded? It is important to strive for consistency here. Similar experiments should result in similar results, if they don't, one has to check the experimental design again for possible error sources and procedural mistakes.
- Analysis and interpretation. How are the results going to be analysed? Merely describing experimental results in words is a possibility, but there are stronger tools available for analysis. Chapter 3 covers a whole range of statistical methods that can be used to detect "significant" effects.
  A very important method used in analysis and interpretation is that of *creating a baseline*. The baseline serves as the fundamental data against which one compares the results obtained in the experiment.
  For example, in work concerned with predictions (of, for example, robot trajectories, temperatures in your home town or stock market developments) one very often compares a prediction method against the baseline of predicting the mean. Predicting that a future signal value will be the mean of past values is a very reasonable prediction, which tends to minimise prediction error. If a prediction method is claimed to be "good", it ought to outperform this simple predictor — something that can be established by the methods described in Chapter 3.
- Often it is useful to conduct a pilot study first, in order to minimise the experimental effort. A pilot study investigates the underlying question in a "broad shot" manner, eliminating certain possibilities, making others more likely, while using simplified and coarser experimental procedures than the eventual final study.

Traps and Countermeasures

Traps

There are a number of known traps to avoid [Barnard et al., 1993]:

1. Confounding effects. If the phenomenon we are interested in is closely correlated with some other effect that is of no interest, special care has to be taken to design the experiment in such a way that only the factor of interest is investigated.

For example, we might be interested in measuring whether the obstacle avoidance movements of a learning mobile robot become more "efficient", smoother, with time. We might find a positive correlation, and conclude that our learning algorithm results in ever smoother movement. But unless we design our experiment carefully, we cannot be sure that the increasingly smooth movement is not the result of decreasing battery charge, resulting in a sluggish response of the robot's motors!

2. Floor and ceiling effects. It is possible that the experimental design is either too demanding or too simple to highlight relevant phenomena.
   For example, we might be interested to investigate whether one service robot performs better than another. If we compare both robots in too simple an environment, they might not show any difference whatsoever (floor effect). On the other hand, if we choose a very complicated environment, neither robot may perform satisfactorily (ceiling effect). Obviously, in order to highlight any differences between the two robots, just the right type of environment complexity is needed.
3. Pseudo-replication (non-independence). The more costly (in terms of time or resources) an experiment, the greater the risk to produce data that is not independent, so-called pseudo-replication. Pseudo-replication means that the errors of our measurements are not unique to each measurement, *i.e.* not independent.
   For example, we might want to measure what effect the colour of objects has on a robot's ability to detect them with its camera system. We could take three different objects, say, and let the robot detect each of these objects ten times. This does not, however, result in thirty independent measurements! We really only have three independent measurements in this case, and need to collapse the ten observations for each object into one value, before we proceed with an analysis of the results.
4. Constant errors, that is systematic errors (biases) can mask true effects, and need to be avoided.
5. "The conspiracy of goodwill" (Peter Medawar). In designing our experiments we need to take great care to retain objectivity. It is very easy to have a particular desired outcome of our experiments in mind, and to research selectively to attain that outcome!

Countermeasures

There are a range of countermeasures that can be taken to avoid the pitfalls just mentioned.

First of all, it is good practice to include *controls* in the experimental design. Such controls can take the form of experiments within the chosen experimental setup whose results are known. Say, for example, an exploration robot is designed to detect certain objects (*e.g.* rocks) in some remote area (*e.g.* Antarctica). The usual procedure, the control, is to test the robot and its ability to detect the objects in a laboratory environment, where the robot's detection ability can be observed and measured.

A second, very commonly used and very effective method to counteract pitfalls of scientific investigation is to work in groups, and to seek independent verification and confirmation of one's experimental setup, experimental procedure, results and their interpretation. Usually hypotheses, setups and interpretations benefit from independent scrutiny!

Constant errors can be avoided by *counterbalancing and randomisation*. Counterbalancing stands for an experimental procedure in which each arrangement of variables under investigation is used an equal number of times. If, for instance, two different robot controllers A and B are to be tested in the same environment, a counterbalanced experimental design would mean that A and B are used first and second respectively for an equal number of time. This would counterbalance constant errors introduced by wear and tear, such as decreasing battery charge.

Another method of dealing with constant errors is that of *randomisation*, by which we mean counterbalancing by chance: the arrangement of variables is determined randomly.

Counterbalancing can only be used if there is no interaction between the counterbalanced variables. If, for example, program B of the above example modified the environment, for instance by rearranging objects in the environment, it does matter in which sequence programs A and B are tested. Counterbalancing would not work in this case.

Dealing with the "conspiracy of goodwill" is relatively easy: a "blind" experimental arrangement will achieve that. *Blind experimentation* means that the experimenter is unaware of the state of the independent variable, and therefore has to log and interpret resulting experimental data at face value, rather than inadvertently putting a slant on the interpretation.

Best known for trials in medicine, where the scientific question is whether a particular drug is effective or not (independent of the patient's and the doctor's knowledge of which drug or placebo is being administered), blind experimentation actually also has a place in robotics. The temptation to interpret results in favour or one's own control program in comparison with a baseline control program is always there! If the experimenter is unaware of which program is currently being run, he cannot possibly log and interpret the data in a biased way!

2.7 Tools Used in this Book

2.7.1 Scilab

In some chapters of this book we have included numerical examples of methods and algorithms discussed in the text. We have used the mathematical programming package `Scilab` [Scilab Consortium, 2004] to illustrate the examples, and included listings of some programs. Many figures in this book were generated using `Scilab`.

`Scilab` is a powerful mathematical programming language, which, as a bonus, has the advantage that it is free for personal use. However, the examples given in this book require few changes to run on other mathematical programming languages, such as for example `Matlab`.

2.8 Summary: The Contrast Between Experimental Mobile Robotics and Scientific Mobile Robotics

In summary, the contrast between mobile robotics as an experimental discipline and mobile robotics as a scientific discipline can be described like this:

- Experimental design and procedure is guided by a testable, falsifiable hypothesis, rather than based on the researcher's personal experience (a "hunch")
- Experimental design and procedure are "question-driven", rather than "application-driven"
- Results are measured and reported quantitatively, rather than qualitatively
- Experimental results are replicated and verified independently (for example by other research groups), rather than presented as stand-alone existence proofs

The following sections of this book will look at how these objectives can be achieved. How can the performance of a mobile robot be assessed, and compared with that of an alternative control program? How can robot-environment interaction be described *quantitatively*? How can testable hypotheses be formulated? How can robot-environment interaction be modelled and simulated accurately? These are the questions that we will investigate now.

Springer
http://www.springer.com/978-1-84628-019-1

Scientific Methods in Mobile Robotics
Quantitative Analysis of Agent Behaviour
Nehmzow, U.
2006, XIV, 210 p., Hardcover
ISBN: 978-1-84628-019-1