Chapter 5
Biorobotics: A Methodological Primer

Edoardo Datteri

Abstract A first objective of this chapter is to present some interesting roles played by biorobotics in the study of intelligent and adaptive animal behaviour. It will be argued that biorobotic experiments can give rise to different "theoretical outcomes", including evaluation of the plausibility of an hypothesis, formulation of new scientific questions, formulation of new hypotheses, support for broad theses about intelligence and cognition, support for broad regulative principles in the study of intelligence and cognition. These outcomes flow from variants of a common procedure, which will be sketched here. A second objective is to introduce some methodological and epistemological problems raised by biorobotics, which will be analysed in reference to the structure of the common procedure, notably concerning the setting-up and execution of "good" experiments and the formulation of "good" explanations of animal behaviour. Knowing and dealing with these problems is crucial to justifying the idea according to which robotic implementation and experimentation can offer interesting theoretical contributions to the study of intelligence and cognition.

Keywords Biorobotic methodology · Simulations · Robotic modelling · Computational neuroethology

5.1 Introduction

The study of intelligence and cognition has been often supported by robots and computing systems. In the early decades of the twentieth century, the building of robotic systems able to interact dynamically and adaptively with realistic environments contributed to promoting a mechanistic, anti-vitalist approach to the explanation of

E. Datteri (✉)
Dipartimento di Scienze Umane per la Formazione "R. Massa", Università degli Studi di Milano-Bicocca, Milan, Italy
e-mail: edoardo.datteri@unimib.it

F. Amigoni and V. Schiaffonati (eds.), *Methods and Experimental Techniques in Computer Engineering*, PoliMI SpringerBriefs,
DOI: 10.1007/978-3-319-00272-9_5,

animal and human behaviours [1]. More recently, robotic reproductions of animal and insect-like behaviours have provided convincing support for the development of general methodological guidelines for the study of intelligence and cognition and, in some cases, have contributed to formulating broad hypotheses on particular aspects of animal behaviour [2]. Robotic simulations have also been used occasionally to test full-fledged models of particular animal behaviours [3]. Bionic systems, connecting robotic components with living biological tissue, have sometimes been deployed for similar theoretical purposes [4].

The term "biorobotics" is often used to indicate the area of research that makes experimental use of robots as outlined above [5]. Similar roles have been assigned to robots and computer systems in the ages of symbolic Artificial Intelligence and Cybernetics. Biorobotic studies had already been carried out before the advent of computers, one of the most cited examples being the "electric dog" built in 1912 and taken by the mechanistic physiologist Jacques Loeb as a test of his theories on phototropic behaviours in moths [1]. The rapid advancement of robotics and computer technologies is now paving the way for fine-grained simulations of sensory-motor biological mechanisms and realistic reproductions of animal shape. However, this technological progress cannot contribute to solving a number of crucial outstanding *methodological* and *epistemological* issues raised by biorobotics. The objective of this chapter is to outline these issues and to convince the reader of their importance: many of the methodological difficulties introduced here, variously concerning the design and execution of "good" biorobotic studies, are not less serious and urgent than the technological difficulties faced when constructing and programming robotic systems. Addressing these issues may contribute to achieving a deeper understanding of the relationship between computing and science, which is one of the main objectives of this book.

The methodological and epistemological problems discussed here may be grouped into two classes.

- **What makes a "good" biorobotic experiment?** A first group of issues is related to the identification of methodological criteria for the design and execution of "good" biorobotic experiments. This group includes issues concerning the relationship between the biorobot and the theoretical model to be tested, the setting-up of an appropriate experimental scenario and, more generally, justification of the inference of theoretical conclusions about the target *biological* system from *robotic* behaviours.
- **What makes a "good" biorobotic explanation?** Most biorobotic experiments are performed to test *explanations* of intelligent behaviours and cognitive capacities. These studies give then rise to the problem of establishing what constitutes a "good" explanation—or more precisely, the problem of identifying criteria to discriminate between "good" explanations and statements that do not deserve this honorific title.

These issues are introduced and discussed in Sects. 5.3 and 5.4, respectively. Section 5.2 provides an overview of biorobotics methodology, illustrating some interesting roles played by biorobotic experiments in the study of intelligence and cognition, and setting the stage for our methodological discussion.

Before proceeding, it is worth stressing that the issues addressed here are biorobotic variants of issues often addressed by philosophers of science in connection with other domains of scientific inquiry. Philosophy of science is chiefly concerned with the rational justification of scientific research methodologies and with the clarification of basic concepts involved in science. The first group of problems addressed here concern the identification of methodological regulative principles for carrying out "good" biorobotic experiments, "good" experiments being those in which one is *justified* in drawing theoretical conclusions (on animal behaviour) from experimental results (i.e., from the analysis of robot behaviour). The second issue concerns the clarification of the notion of "explanation" in biorobotics. These philosophical issues are of fundamental importance for biorobotics researchers, as the ensuing discussion will show.

5.2 On Various Experimental Roles of Biorobotics

Section 5.2.1 provides an overview of biorobotics methodology, while Sect. 5.2.2 illustrates some interesting roles played by biorobotic experiments in the study of intelligence and cognition.

5.2.1 The Methodological Core

As pointed out in Sect. 5.1, biorobotics can contribute in various ways to the study of intelligence and cognition. These theoretical contributions flow from variants of a common experimental procedure, which is schematically described here with reference to a purely notional biorobotic case study. Figure 5.1 will help follow the various steps of the methodology.

Any given biorobotic study will focus on a particular class L of biological systems (see the top-right box in Fig. 5.1). Suppose L is the class of rats. Typically, biorobotic studies are carried out to explore the mechanisms underlying the manifestation of particular *capacities* CL possessed by members of class L and exhibited by them in particular conditions EL (middle-right box of Fig. 5.1). For example, many biorobotic studies focus on the capacity of rats to orient themselves in experimental mazes. What is a "capacity"? Many philosophers have tried to clarify this notion [6]. Here, consistently with [7], the term will be taken to refer to a (behavioural) *regularity* expressed by a *generalization* statement. For example, rats' capacity to orient themselves in mazes may be expressed as the generalization according to which rats travel from an initial position to a destination point in a maze, making fewer errors on each new trial than on previous trials [8].

Target capacity CL is explained by a hypothetical mechanism description ML (bottom right box of Fig. 5.1). Most mechanism descriptions formulated in the cognitive sciences make reference to a number of interacting *components* within system L, each one playing a distinct role in the exhibition of the target capacity. The notion

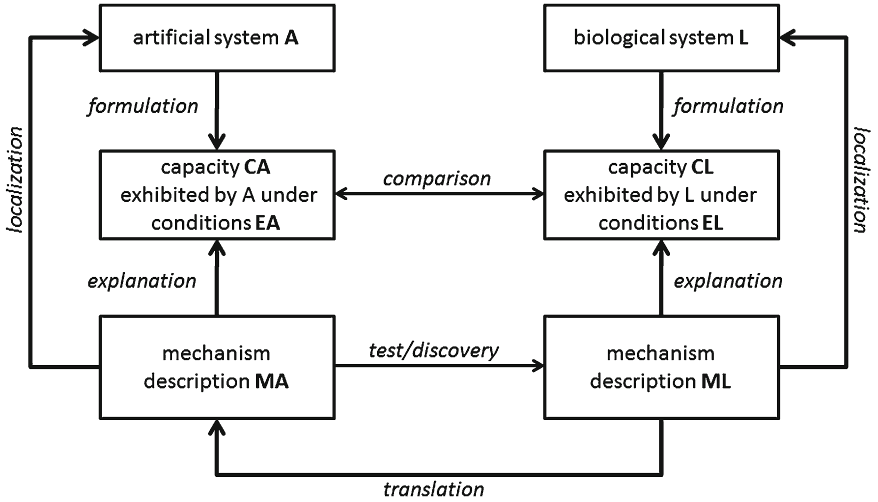

Fig. 5.1 A sketch of biorobotic methodology

of “mechanism” is intrinsically connected with the notion of “regularity”: individual components of a mechanism exhibit regular behaviours and interact with each other in regular ways. This is the reason for which cognitive science mechanisms are described by sets of *generalizations*, i.e., of statements expressing regularities in the behaviour of components and their interactions. For example, rat orientation capacities are often explained in terms of a mechanism including a number of sensory and motor organs of the rat and a number of brain areas, notably including the area of the hippocampus containing the so-called *place cells*. The behaviour of these components is assumed to be regular: for example, the firing of each place cell is believed to occur only when the rat is in a particular spatial position, under a number of reasonably well-known boundary conditions [9]. Note that this is a neuroscientific example. Cognitive science generalizations may be couched in non-neuroscientific vocabulary, e.g., vocabulary making reference to mental states such as beliefs, desires and intentions, or to the physical properties of the system, including its morphology [2].

How may we test the hypothesis according to which ML is the mechanism enabling L to exhibit CL? Here is the biorobotic answer: build an artificial (robotic) system A whose behaviour is governed by the mechanism ML under scrutiny, put A under environmental conditions EL, and observe its behaviour (see the left column of Fig. 5.1). This amounts to building a robotic *simulation* of ML and to experimenting on it. If A exhibits capacity CL (i.e., if the robotic simulation reproduces the target biological behaviours), one may be induced to conclude that ML is the mechanism enabling L to exhibit CL. A different result may be taken to support rejection of ML. To be sure, the artificial system A cannot be said to implement ML (which will probably involve *biological* components, such as brain tissues) but a sort of “artificial translation" MA of ML (which will make reference to *artificial* components, such as electronic circuit boards, fulfilling similar functional roles). And the experiments are intended to check, by behavioural comparison, whether the capacity CA exhibited by the artificial system in an experimental environment EA is similar, in a sense to be clarified, to the target capacity CL (see the "comparison" arrow in Fig. 5.1). The "discovery" and the left "explanation" arrows will be discussed later.

5.2.2 Variants and Applications

Experiments with A may play different roles in the study of capacity CL. In particular they can give rise to different "theoretical outcomes", including evaluations of the plausibility of a hypothesis, the formulation of new scientific questions, the formulation of new hypotheses or refinements of previous ones, support for broad theses about intelligence and cognition, and support for broad regulative principles in the study of intelligence and cognition (but the list is not exhaustive). These outcomes flow from variants of the procedure illustrated above, and different outcomes are often obtained at different steps of the same study. Although there may be some overlap, it is worth making these distinctions in order to fully appreciate the value of biorobotics experimentation in the cognitive sciences.

5.2.2.1 Corroboration of Mature Hypotheses

A mechanistic hypothesis ML may be more or less "mature", in the sense of having received more or less support from previous experiments or auxiliary assumptions. In particular, some biorobotic hypotheses are backed by well-corroborated *localization assumptions*, i.e., assumptions stating that some aspects of the mechanism are actually localized somewhere in the target system L. For example, [10] carried out a biorobotic study on a mechanistic hypothesis regarding hippocampal *place cells*, which behaved as prescribed by well-corroborated neuroscientific hypotheses. A large number of features of the mechanistic hypothesis tested in [11] were assumed to be localized in the rat brain and to behave as prescribed by the hypothesis. Mature mechanism descriptions have been tested in biorobotic studies on ant self-localization [12], on cricket phonotaxis [13], on motion detection in locusts [14].

If the hypothesis is already mature in the sense assumed here, why should one perform biorobotic experiments on it? This question points to the unique experimental value of biorobotics (and of simulative methodologies at large). Many conventional experimental methodologies used in the cognitive (neuro)sciences may enable one to evaluate whether aspects of ML are localized in L, e.g., if L has the components mentioned in ML, if these components behave as prescribed by ML, if they are connected as specified by ML. What is missing is the "bridge" between these localization assumptions and the capacity under investigation. Having good reasons to believe that many aspects of the mechanism ML are localized in L is of course crucial to the adequacy of the explanation, but does not amount to claiming that ML *is the mechanism underlying CL*. Some of these aspects could be unrelated to CL. Conversely, some aspects essential to CL could be missing in ML. The hypothesis could well make reference to the "right" components, but connect them in a way that is not functional to CL. Here we may acknowledge one of the distinctive experimental roles of biorobotics (and of computer simulation) in the study of intelligent behaviour: a biorobotic experiment on ML can contribute to assessing whether *these* components, organized in *this* way, produce the behaviour CL of interest. In more general terms, the biorobot crucially contributes to *evaluating the plausibility of a mechanistic hypothesis* ML regarding CL.

Needless to say, maturity comes by degrees: even very mature biorobotic hypotheses may still have aspects that are not backed by well-corroborated localization assumptions. We will discuss this case in Sect. 5.2.2.3.

5.2.2.2 Formulation of New Scientific Questions

Mismatches between CL and CA typically stimulate the formulation of new questions calling for an explanation of the result. For example, the robotic lobster described in [15] consistently failed to match biological performance in a chemiotaxis task under particular conditions. Why? The reason could be sought in biologically unmotivated aspects of the robot (e.g., battery discharge) or in the hypothesized mechanism description. Eventually the authors concluded that robot failures were due to limitations of the proposed hypothesis ML. Biorobotic experiments, in this case, stimulated the *formulation of a new scientific question* which was followed by the formulation of a new mechanistic hypothesis. Note that this question concerns whether the mechanism description ML produces the target behaviour CL or not. In the previous section we noted that biorobotics can play a crucial role in *addressing* questions concerning the "bridge" between mechanisms and their behaviours. Here we add that biorobotics can also play a unique role in *raising* questions of this kind.

Biorobotic experiments may also enable unexpected behaviours of L to be identified, stimulating new why-questions regarding them. The previous example was a case of falsification; now suppose, on the contrary, that A matches L's behaviour to a great extent in a variety of experimental conditions. As discussed so far, this result may be taken to corroborate the hypothesis according to which ML produces CL. Now suppose that A is put in novel internal or environmental conditions EA' producing peculiar behaviours that have never been observed in L (possibly because L's behaviour has never been observed under conditions EA'). Will L display the same peculiar behaviours in EA'? Here, new scientific curiosity regarding L has been raised by experiments with a robotic surrogate. It goes without saying that the same question could have been raised without any robotic experimentation. However, in many cases, manipulation of the robot may be more feasible for practical or ethical reasons than manipulation of the target biological system. And the initial corroboration may support the hypothesis that A and L share the same mechanism, thus increasing the expectation that L will display the peculiar behaviours in EA’ and providing strong motivation for addressing the new question.

5.2.2.3 Formulation of New Hypotheses

Let us start from a notional example. Contrary to the cases described so far, suppose that no plausible mechanistic hypothesis on CL is available to fill the “ML” box in Fig. 5.1. This may be due to the fact that previous hypotheses have been discarded, or to the fact that capacity CL has been newly observed. Nonetheless, it is possible to build a robot A whose behaviour CA is similar to CL, by following a conventional, iterative process of robotic design, implementation and testing. If A reproduces the behaviour of interest, one may be induced to “translate” aspects of the mechanism MA implemented in the machine into a new mechanistic hypothesis ML for L, which simultaneously receives initial biorobotic support. This is the “discovery” arrow in Fig. 5.1, the formulation of a new hypothesis being one of the types of discovery taking place in science. In short: if the robot behaves like L, one may be legitimately induced to hypothesize that L produces CL by virtue of the same mechanism implemented in the machine, thus filling the “ML” box. In this case, the role of the biorobotic experiment is to *support the formulation of a new hypothesis regarding CL*. Here is a concrete example. Many robots, built for purposes that are totally unrelated to biological research, produce avoidance and attraction behaviours by virtue of crossed excitatory or inhibitory connections between sensors and motors, as in Braitenberg’s vehicles 2a and 2b [16]. These purely robotic implementation successes have stimulated the formulation of a hypothesis about lobster chemiotaxis based on a similar mechanism [15].

It is worth noting that this example and those discussed in the previous sections share important aspects of the methodological procedure described in Sect. 5.2.1: in both cases, behavioural comparisons between A and L provide elements for reflection on the similarity between MA and ML. The two cases differ with regard to the maturity of the hypothesis. A relatively mature hypothesis was available before robotic implementation in the examples discussed in Sect. 5.2.2.1 while in case just outlined no previous hypothesis is available: it is newly formulated via translation from MA.

Processes of translation from robotic mechanism descriptions to biological hypotheses often take place in biorobotics, even when a biological hypothesis ML is available. MA may have features that are not reflected in ML—as we will discuss in Sect. 5.3, all biorobots include aspects that are not mentioned in the biological hypothesis under scrutiny. In particular, MA may include *components* that are necessary for producing the desired behaviour even though they are not mentioned in ML (e.g., because scientific theorizing on CL is still in its infancy). In this case, success on the part of A in replicating the behaviours of L may induce one to include those components in ML as well. For example, the robot described in [10] included so-called artificial *goal cells* which were necessary to memorize goal locations, and therefore to build a system able to fully replicate the maze navigation capabilities observed in rats. In the experiments the robot displayed goal-seeking abilities, and this was taken to support the hypothesis that something functionally equivalent to goal cells could be found somewhere in the rat brain: the experiment supported the *formulation of a new localization assumption* to the effect that goal cells can be found in the rat brain.

In other cases a *mechanistic hypothesis*, and not simply a localization assumption, is obtained by translation from MA. In the study on cricket phonotaxis described in [13], the robot was found to be affected by environmental conditions that do not affect "real" cricket behaviour: the grass on which both systems were placed slowed down the robot. This was not only due to the particular shape of the robot, but also—as argued by the authors—to the fact that it moved without any feedback-based control mechanism able to correct deviations due to the grass. An obvious solution to this robotic problem would be to provide the system with a feedback-based control mechanism of that sort. This suggestion gives naturally rise to a new mechanistic hypothesis about crickets: they are likely to have some form of feedback-based mechanism to overcome the friction caused by the grass and move efficiently on it.

Finally, *refinements of previous hypotheses* on L are very often obtained by translation from MA. This typically happens when the initial hypothesis under scrutiny ML is formulated in vague terms, e.g., when it includes unfixed parameters that must be fixed in order to obtain a working system [17]. For example, the mechanism description ML tested in the aforementioned study on lobster chemiotaxis [15] included no prescription regarding the distance between the two chemical sensors to be put at each side of the robot. Needless to say, this parameter must be fixed in order to build the robot—the sensors must be put at *some* distance! Let us call ML' the mechanism description ML with distance value fixed. As in the previous examples, ML' is obtained by "translation" from MA: the exact distance value is specified in a description MA of the mechanism implemented in the robot. ML and ML' are clearly different in at least one respect—not a trivial one indeed, as intra-sensor distance *may* actually matter for a robot whose behaviour is dependent on the difference between the stimuli perceived at each of two sensors. And one may legitimately claim that the hypothesis actually tested in the study is the *latter* one: strictly speaking the robot must be considered as a simulation of the *refined* mechanism description ML' rather than of the vaguer hypothesis ML. This example will be discussed again in Sect. 5.3 in connection with the issue of evaluating simulation accuracy.

5.2.2.4 Constructive Proof of Mechanicism

We have argued that biorobotic experiments can contribute to formulating novel mechanistic hypotheses about the behaviour of living systems and their cognitive capacities. In some cases, especially in the early decades of the twentieth century, this has amounted to supporting *mechanicism*, understood as an epistemological orientation towards explaining events by identifying the mechanism producing them. *Vitalist* philosophers and physiologists, including Henri Bergson and Hans Driesch, believed that many aspects of intelligence and cognition, including learning, could be explained only by appeal to *non-mechanical* vital forces. Let us call CL a general description of a particular cognitive or behavioural capacity. In many cases, the construction of a robot exhibiting a behaviour (CA) substantially similar to CL has provided support for the thesis that CL *may be explained mechanistically*: insofar as at least *one* mechanistic explanation of CL had been found to exist, namely the mechanism implemented in the machine A (in our terms, the mechanism ML obtained by translation from MA). Many such cases are discussed in [1], including Ashby's homeostat, Hull's 'psychic machine', Grey Walter's tortoises and, in more recent times, Braitenberg vehicles (see also [18]).

5.2.2.5 Guidelines for Explaining Intelligence

The so-called 'embodied approach' in Cognitive Science and Artificial Intelligence is based on a number of broad theses concerning the nature of intelligent behaviour and the appropriate way to explain it. The most basic of these propositions will by now seem obvious to most of us: the behaviour of a living system is not only determined by the control mechanism implemented in it but also by its interaction with the external world—and a "simple" control mechanism can produce "complex" behaviours due to the "complexity" of the environment, as stressed in [16, 19], and others. This claim gives rise to a variety of broad theses regarding the most appropriate way to build efficient robots and explain intelligent behaviour. As far as explanation is concerned, it is stressed that in explaining the behaviour of a system, particular attention should be paid to its shape and to the features of its *ecological niche*. This may help to avoid what Braitenberg believed to be a bias typical of (cognitive) scientists, that is to say, the tendency to explain "complex" behaviours by appeal to "complex" mechanisms.

This *guideline for explaining intelligence and cognition* has been supported by the implementation of robots able to exploit their shape and physical dynamics, rather than sophisticated control mechanisms, to generate apparently "complex" behaviours. Cases in point are the Swiss and Stumpy robots, developed at the Artificial Intelligence Laboratory of the University of Zurich [2]. The robot-based methodology supporting this guideline for explanation fits well with the procedure described in Sect. 5.2.1, and the case we are discussing is substantially similar to those discussed in Sect. 5.2.2.3. Robot A generates behaviours that are very similar to those observed in a broad class L of living systems. The mechanism used, MA, exploits particular features of A's shape and environment. Similarly to the case discussed in Sect. 5.2.2.3, this result may be taken to support the broad claim that L's behaviour may be explained by a sort of "biological translation" of MA, i.e., that L's behaviour can be explained by appeal to particular features of L's shape and environment. As a result, the robot may be regarded as a positive implementation of the guideline for explaining intelligence outlined above, according to which particular attention should be paid to the shape of a system and to the features of its ecological niche in explaining its behaviour.

5.3 What Makes a Good Biorobotic Experiment?

We have discussed various ways in which biorobotics can contribute to the study of intelligence and cognition. The procedures leading to these results share the common methodological structure illustrated in Sect. 5.2.1. And they also share a number of epistemological and methodological problems affecting that methodological structure. Figure 5.1 may help to identify these problems, some of which are related to the design and execution of a "good" biorobotic experiment.

5.3.1 Experimental Comparisons Between CA and CL

Webb [20] has convincingly argued that every biorobotic inquiry regarding intelligence and cognition *in living systems* must be based on some kind of comparison between biological and robotic behaviours. No interesting insight into animal behaviour can be obtained by reasoning solely about robot behaviours, contrary to what has been suggested by proponents of the so-called "animat" approach within robotics. In "good" biorobotic experiments, one draws theoretical conclusions about ML from the result of experimental comparisons between CA and CL. This immediately gives rise to a methodological justification problem: how should these comparisons be carried out for their results to play a legitimate role in the testing and discovery of ML?

In particular, what aspects of the two behaviours should be considered in the comparison? In the aforementioned study on lobster chemiotaxis, for example, the authors focused on the success rates of robotic and "biological" lobsters in reaching the destination site, irrespectively of the trajectories followed by the two systems. Finer-grained comparisons of robotic and human elbow trajectories were made in the biorobotic study on forearm posture maintenance described in [21]. Clearly, the outcome of the comparison between CA and CL (and, consequently, the outcome of the whole biorobotic study) crucially depends on, amongst other factors, the particular aspects being compared: the robot may match L's rate of success in reaching the destination by following completely different trajectories. How may "good" matching criteria be chosen?

This is by no means an easy methodological question, and here we can only provide some prompts for further discussion. In principle, the range of possible matching criteria is very wide in any biorobotic study. However, it is reasonable to claim that what constitutes the "right" criteria *depends on the scientific question* addressed in the study. If one aims to explain why lobsters' rate of success in reaching the source of a chemical stream is so high, then one should look at the rate of success of the robotic simulation. If, instead, one aims to explain why lobsters generate certain specific trajectories as opposed to others, then one should compare robot and animal trajectories. Conversely, comparisons between rates of success will legitimately enable one, at most, to theorize on the animal success rate and not, in principle, on the trajectories generated by the living system.

A related crucial methodological question concerns the setting-up of the experimental setting EA in which to observe robot behaviours. Ideally, in most cases, robots can operate in environmental conditions that are very similar or identical to the animal's ecological niche: an example is the robot used to study the localization abilities of the Cataglyphis desert ant, which is used in the Sahara desert [12]. However, in many studies, robot and biological behaviours are observed in quite different environments [10]. If EA and EL are substantially different, is one justified in taking A's behaviours as empirical evidence in reasoning about the mechanism producing CL in EL? In principle, the analysis of A's behaviour in EA could enable one, at most, to theorize on the behaviour produced by L in EA. However, on closer scrutiny, this seems to be too strong a position: one may reasonably claim that some degree of resemblance between EA and EL licenses *some kind* of theoretical conclusion regarding L's behaviour in EL. This methodological justification problem calls for the identification of regulative principles governing the set up of "good" experimental environments in biorobotics. To address this problem it is worth stressing that why-questions investigated by the cognitive sciences typically do not concern animal behaviours observed in their ecological niche, but rather behaviours observed in specifically tailored and controlled experimental settings. Robots can go wild [22], but biorobotic experiments are often designed to test hypotheses on animal behaviours observed in laboratory settings [23].

5.3.2 Simulation Accuracy

Not *every* robot can contribute to testing a mechanistic hypothesis ML. Needless to say, a commercial robotic vacuum cleaner can hardly provide empirical evidence to test a hypothesis on human posture maintenance. There must be a close relationship between ML, MA and A in order to make legitimate use of A in the testing of ML. In particular, it has been often claimed that the robot itself should be a *good simulation of the hypothesis ML under scrutiny*. Otherwise, it is not clear why robotic behaviours should be taken as empirical evidence in reasoning about ML. But what is the nature of this close relationship? What makes a good robotic simulation of a mechanistic hypothesis? In other words, what kind of criteria should be used to check if A is a good simulation of ML?

According to a plausible interpretation of the term "simulation", A is a good simulation of ML if A works *as prescribed by ML* or, equivalently, if A implements the mechanism described in ML. However, on closer scrutiny, this condition seems hard to attain. First, as already noted in Sect. 2.2.3, cognitive science mechanistic hypotheses are often vaguely and qualitatively specified; any robotic implementation of them may be regarded, at least in principle, as the implementation of a *fully specified* version of the initial hypothesis. Second, all robots will include components that are not mentioned in ML: animals do not need DC batteries. Third, the process of robotic implementation often involves approximations and adjustments with respect to the initial mechanism. For these and other reasons (more extensively discussed in [17] and [3]), it seems unlikely that the mechanism MA actually implemented in the machine will be *exactly the same as* the mechanism ML under scrutiny. In this case, are there rational grounds for justifying the use of A to reason about ML?

This question is still open and is not easily answered. Indeed, in most biorobotic studies, it is simply claimed that the robot "is closely based on", "implements accurately", "simulates in detail" the target hypothesis, but these claims are not well clarified and justified. And in many cases, closer scrutiny will identify non-trivial discrepancies between the hypothesis and the implemented mechanism. The problem of defining criteria for evaluating simulation accuracy has been occasionally addressed in the methodological literature (e.g., [24]) but a satisfactory solution is still lacking. Without purporting to solve the issue here, let us briefly propose a means of viewing this problem from another, possibly more fruitful, perspective.

We have reasoned about the possibility of building a robot A that behaves *exactly like* the hypothesis ML under scrutiny. And we have pointed out that every biorobot A possesses some features that are not specified by the biological hypothesis and are therefore determined on the basis of other criteria. However, we should avoid jumping too quickly from the existence of these differences to the conclusion that A is a bad experimental tool for testing ML. This point can be aptly illustrated by reference to the experiment on lobsters chemiotaxis described in [15]. As outlined above, ML contained no prescription regarding the distance between the chemical sensors: it was a vague hypothesis at least in this respect. This ambiguity had to be addressed in order to build A, and the fixed intra-sensor distance value was specified in a description MA* of A (the reason for the asterisk will become clear later). So, there was at least one difference between MA* and ML. However, the experiments showed that distance value was totally *irrelevant* to the robot's ability to replicate the behaviour of interest. Indeed, the authors conducted several experiments with different intra-sensor distances, finding that the level of behavioural match between biological and artificial behaviours did not change. Therefore the difference between MA* and ML was irrelevant with respect to the outcomes of the behavioural comparisons between the two systems—the particular intra-sensor distance chosen by the authors *did not make the difference with respect to whether A replicated L's behaviours or not*, in the sense that different intra-sensor distances did not give rise to different behaviours (i.e., they did not increase the robot's success rate). According to many plausible accounts of what it is to explain something (see, e.g., [25]), we therefore would not mention intra-sensor distance in a mechanistic explanation of A's ability to reproduce L's behaviours. For similar reasons, the fact that A's external structure is made of polyethylene instead of polypropylene would not be mentioned in a mechanistic explanation of A's behaviours, provided that this choice does not make any relevant behavioural difference. Let us call MA this mechanistic explanation, which is silent on intra-sensor distance. Now, it seems reasonable to claim that MA is superior to MA* in describing *the* mechanism producing CA: MA and not MA* includes a specification of the aspects that are actually *relevant* to A's behaviour. So, why worry about the difference between ML and MA*? It is reasonable to claim that what is important is whether there are differences between ML and MA, i.e., whether there are differences between ML *and the mechanisms that are really relevant to the manifestation of CA*.

The conclusion of this reflection is that not *every* difference between the hypothesis under investigation and the robot is relevant to whether the latter is a good tool to reason about the former. What matters is whether the biological hypothesis is similar, in a sense to be specified, to the mechanism MA which has actually governed the robot in the experiments; and, for the reasons discussed here, not every peculiarity of the robot needs to be specified in MA. In this perspective, evaluation of simulation accuracy crucially requires the formulation of good explanations of *robotic* behaviours. This claim, which is left to the reader as an insight for further analysis, introduces us to another fundamental methodological issue concerning biorobotics.

5.4 What Makes a Good Biorobotic Explanation?

Biorobotic studies are typically aimed at formulating *good explanations* of biological behaviours (see the arrow labelled "explanation" connecting ML and CL in Fig. 5.1). However, no clear and precise criteria are available to distinguish "good explanations" of a given event or regularity from statements that do not deserve this title. Many substantially different mechanistic hypotheses may be formulated to explain capacity CL—how should we choose among them?

A bioroboticist may suggest using a biorobotic test: if a robotic simulation of ML reproduces the target behaviour, then accept ML as a good basis for explaining CL. However, this is only part of the story. Simulation success may corroborate ML but it seems to be *insufficient* to conclusively claim explanation adequacy. The feedback-based hypothesis ML about lobster chemiotaxis described above prescribes that each chemical sensor is positively connected to the motor organs located on the opposite side of the system. Suppose that a robotic simulation of this hypothesis performs efficient chemiotaxis. This result may induce one to corroborate ML, but will not dispel all doubts about its explanatory adequacy. First, ML is very "simple": no mention is made of the mechanisms actually connecting sensors with motor organs (there is surely more than a pair of excitatory neurons) or of the gait control mechanisms. Second, ML is very "idealized": the hypothesis only makes sense if we assume that no external or internal perturbation will affect lobster behaviour (no mechanism is included to resist water turbulence, avoid predators, choose between competing internal motivations, and so on). One may claim that these are not serious objections and that ML is a good basis for explanation even though it is simple and idealized. However, this claim needs justification—and it cannot be adequately justified without appealing to some notion of what makes a good explanation!

This is clearly a crucial and urgent methodological issue for anyone who aims to *explain* intelligent animal behaviours. And it is still an open issue, similarly to those discussed in the previous sections. A thorough discussion is beyond the scope of this chapter. Our aim is merely to provide some insights for discussion, by providing a brief overview of the two main positions put forward in the philosophical literature regarding this problem.

Let CL refer to the statement expressing what is to be explained, for example, "rats travel from an initial position to a destination point in a maze, making fewer errors on each new trial than on previous trials". And let "Exp" refer to the statements constituting the proposed explanation, for example, statements describing a neural mechanism *plus* other statements specifying some initial or boundary conditions. According to the so-called *ontic* view of scientific explanation, Exp is a good explanation of CL if and only if a particular relationship—i.e., a *causal relationship*—holds between what is described by Exp and what is described by CL. In short, a good explanation of something describes its causes. This is a very plausible and commonsensical position ("explaining" is often used as synonymous of "finding the causes" in the everyday language). However, it is seriously affected by the difficulty encountered in defining the notion of a "cause"—or more precisely, by the lack of criteria for distinguishing causal relationships from non-causal generalizations [26].

An alternative position on the nature of scientific explanation is the so-called *epistemic* view. The idea is that Exp is a good explanation of CL if and only if a particular relationship holds between *knowledge of* what is described by Exp and *knowledge of* what is described by CL (the italics mark the difference with respect to the ontic thesis). In particular, many supporters of the epistemic approach claim that, in a good explanation, knowledge of what is described by Exp should have allowed one, if taken into account in time, to *predict* what is described by CL (prediction being a sort of "knowing in advance"). This is a plausible idea, at least in general terms. The feeling of having received a good explanation of an event is very often accompanied by the feeling that, had we known the explanation in time, we could have predicted that event. This view is not affected by the many problems arising from the notion of "causation", simply because it does not include that notion. However, particular epistemic models of scientific explanation—notably the so-called deductive-nomological model [27]—have been widely criticized in the philosophical literature [26].

As far as *biorobotic* explanations are concerned, there are good reasons to claim that they presuppose an epistemic account of scientific explanation according to which the ability to predict the behaviour of interest is at least *essential* (though not sufficient) for a good explanation. The main reason is that, in this methodology, robotic systems are used to identify implications—which in many cases consist of *predictions*—of the explanatory hypothesis ML under scrutiny. This claim needs further clarification and justification, which is beyond the scope of this brief introduction to the philosophical problem of scientific explanation. As previously noted, the problem is still open: careful analyses of "good" and "non-good" scientific explanations, possibly drawing on the biorobotic literature, are needed to solve it.

5.5 Summary and Conclusion

A first objective of this chapter was to present some interesting roles played by biorobotics in the study of intelligent and adaptive animal behaviour. We have claimed that biorobotic experiments can give rise to different "theoretical outcomes", including evaluation of the plausibility of an hypothesis, formulation of new scientific questions, formulation of new hypotheses, support for broad theses about intelligence and cognition, support for broad regulative principles in the study of intelligence and cognition. We have shown that these outcomes flow from variants of a common procedure. A second objective was to introduce some methodological and epistemological problems raised by biorobotics, which we have analysed in reference to the structure of the common procedure, notably concerning the setting-up and execution of "good" experiments and the formulation of "good" explanations of animal behaviour. Knowing and dealing with these problems is crucial to justifying the idea according to which robotic implementation and experimentation can offer interesting theoretical contributions to the study of intelligence and cognition, and may contribute to achieving a deeper understanding of the relationship between computing and science, which is one of the main objectives of this book.

References

1. Cordeschi R (2002) The discovery of the artificial: behavior, mind and machines before and beyond cybernetics. Kluwer Academic Publishers, Dordrecht
2. Pfeifer R, Bongard J (2007) How the body shapes the way we think: a new view of intelligence. The MIT Press, Cambridge (MA)
3. Datteri E, Tamburrini G (2007) Biorobotic experiments for the discovery of biological mechanisms. Phil Sci 74(3):409–430
4. Datteri E (2009) Simulation experiments in bionics: a regulative methodological perspective. Biol Phil 24(3):301–324
5. Webb B (2001) Can robots make good models of biological behaviour? Behav Brain Sci 24:1033–1050
6. Cartwright N (1994) Nature's capacities and their measurement. Oxford University Press, Oxford
7. Cummins R (1985) The nature of psychological explanation. The MIT Press, Cambridge (MA)
8. Tolman E (1948) Cognitive maps in rats and men. Psychol Rev 55(4):189–208
9. Moser E, Kropff E, Moser M (2008) Place cells, grid cells, and the brain's spatial representation system. Annu Rev Neurosci 31:69–89
10. Burgess N, Jackson A, Hartley T, O'Keefe J (2000) Predictions derived from modelling the hippocampal role in navigation. Biol Cybern 83(3):301–312
11. Krichmar J, Seth A, Nitz D, Fleischer J, Edelman G (2005) Spatial navigation and causal analysis in a brain-based device modeling cortical-ippocampal interactions. Neuroinformatics 3(3):197–222
12. Lambrinos D, Möller R, Labhart T, Pfeifer R, Wehner R (2000) A mobile robot employing insect strategies for navigation. Robot Auton Syst 30:39–64
13. Reeve R, Webb B, Horchler A, Indiveri G, Quinn R (2005) New technologies for testing a model of cricket phonotaxis on an outdoor robot. Robot Auton Syst 51(1):41–54

14. Blanchard M, Rind F, Verschure P (2000) Collision avoidance using a model of the locust LGMD neuron. Robot Auton Syst 30(1–2):17–38
15. Grasso F, Consi T, Mountain D, Atema J (2000) Biomimetic robot lobster performs chemo-orientation in turbulence using a pair of spatially separated sensors: progress and challenges. Robot Auton Syst 30(1–2):115–131
16. Braitenberg V (1986) Vehicles: experiments in synthetic psychology. The MIT Press, Cambridge (MA)
17. Tamburrini G, Datteri E (2005) Machine experiments and theoretical modelling: from cybernetic methodology to neuro-robotics. Mind Mach 15(3–4):335–358
18. Schlimm D (2009) Learning from the existence of models: on psychic machines, tortoises, and computer simulations. Synthese 169:521–538
19. Simon H (1969) The sciences of the artificial. The MIT Press, Cambridge (MA)
20. Webb B (2009) Animals versus animats: or why not model the real iguana? Adapt Behav 17(4):269–286
21. Chou P, Hannaford B (1997) Study of human forearm posture maintenance with a physiologically based robotic arm and spinal level neural controller. Biol Cybern 76(4):285–298
22. Knight J (2005) Animal behaviour: when robots go wild. Nature 434:954–955
23. Datteri E, Laudisa F (2012) Model testing, prediction and experimental protocols in neuroscience: a case study. Stud Hist Phil Biol Biomed Sci 43(3):602–610
24. Webb B (2006) Validating biorobotic models. J Neural Eng 3(3):R25–R35
25. Woodward J (2003) Making things happen: a theory of causal explanation. Oxford University Press, New York
26. Psillos S (2002) Causation and explanation. Acumen - McGill Queens University Press, Georgetown, Canada
27. Hempel C, Oppenheim P (1948) Studies in the logic of explanation. Phil Sci 15(2):135–175