Evolution: Education and Outreach
a SpringerOpen Journal

BOOK REVIEW

Open Access

CrossMark

Telescopes on the past: robots as models of extinct animals

Edoardo Datteri*

Book details

Darwin's Devices. What Evolving Robots Can Teach us About the History of Life and the Future of Technology, by John Long. New York: Basic Books, 2012. pp. 273. H/b $ 26.99 US, $30.00 CAN.

Why do animals behave as they do? In *The Study of Instinct* (1951), Nikolaas Tinbergen famously distinguished between four different ways of interpreting this question. A first potential interpretation frames the question as pointing to the sensory-motor *mechanisms* governing animal behavior (what mechanisms determine changes in the skin color of chameleons in particular circumstances?). However, the same why-question could be also interpreted as pointing to the causal mechanisms governing the *development* of those sensory-motor mechanisms from the birth of the animal onwards (why it is that chameleons develop those color-changing mechanisms during their ontogenesis?). A third possible interpretation is *evolutionary*: how it is that those particular mechanisms have evolved in a particular animal during the history of the species? Finally, one may adopt a *teleological* interpretation of the question ("why do chameleons change colors?" would thus be interpreted as "what is the color-changing mechanism for?"). Note that the first three types of question, but not the fourth, are appropriately answered by describing *mechanisms* (sensory-motor, ontogenetic, and evolutionary, respectively). In contrast, the fourth type does not invoke a mechanism of any kind, but enquires into the *purpose* of the behavior produced by the mechanism of interest.

Robotic and computer systems have been used since the beginning of twentieth century to address Tinbergen's first class of why-question (see *The Discovery of the Artificial* by the late Roberto Cordeschi, Springer 2002, for a

discussion). To understand whether a particular sensory-motor mechanism is truly responsible for a particular form of behavior, one may simulate the mechanism in a machine and assess whether it is capable of reproducing the target behavior. This strategy characterizes a significant proportion of studies in Artificial Intelligence, and underpins what we know as contemporary biorobotics. In his fascinating "Darwin's Devices" John Long, Professor of Biology and Cognitive Science at Vassar College, explains how they may also be used to address the third kind of question in Tinbergen's taxonomy, namely questions about the evolution of particular traits during the history of a species.

Long's book provides a detailed treatment of the methods of this *evolutionary* branch of *biorobotics* and may therefore be particularly informative and stimulating for roboticists and biologists who have never reflected on how robots might be used for such a purpose. The author outlines and extensively discusses the key steps in these methodologies, including formulation of the target research question, the setting-up and performing of experiments (which characteristically involve constructing robots as models of extinct organisms), interpretation of experimental results, and evaluation of the initial hypothesis in light of the experimental outcomes. Long also dwells at length on the relative advantages of using robotic models to study evolution as opposed to more traditional experimental techniques. Furthermore, I believe that the book may also be usefully read by those who do not have direct experience of scientific research. For example, a key part of the work is devoted to the discussion of what is technically known as a case of falsification: Long's experiments with evolving robots ultimately

*Correspondence: edoardo.datteri@unimib.it
"R. Massa" Department of Educational Human Sciences, University of Milano-Bicocca, Milan, Italy

Springer

produced results that were at odds with the predictions of the research team. His analysis of how the research group reacted to the initial disappointment of obtaining apparently incongruent data points up key transversal aspects of doing science, that also apply outside the field of evolutionary robotics. Reading the book will thus provide those who are unfamiliar with scientific research with an appreciation of the multiple theoretical complexities faced by scientists in the course of their discovery processes.

Let me begin from the first point. Long's book helps us to understand how robots may be used to study evolution. To this end, he outlines two kinds of experimental strategy involving robots as models of extinct organisms, which may be adopted to address the third kind of question in Tinbergen's taxonomy. More specifically, Long claims that robots may be used to address both sides of what he considers to be the fundamental question to be asked about any evolutionary system: why have some traits evolved (Side 1) but others not (Side 2)? Now, Long's own specific interest concerns the evolution, in fish, of vertebrae from notochords, or continuous collagenous rods without bones. So the key questions addressed by his group are: Why have vertebrae evolved from notochords? And why it is that other structures have not evolved?

In relation to the first of these questions, Long's group formulated various hypotheses, which are extensively reviewed in the book. By way of illustration, I next discuss one of them in more detail (see Chapter 3), namely the hypothesis that vertebral columns evolved from notochords because they enable individuals to swim faster, and therefore to locate and eat a greater amount of found, thereby enhancing their chances of survival and reproductive success. Long's approach to testing this hypothesis comprised the following steps: (1) he and his team built populations of light-seeking swimming robots (labelled Tadro3) that varied in terms of the stiffness of their notochord; each robot had a simulated genetic profile coding its notochord stiffness; (2) the robots were allowed to swim in an experimental water arena fitted with a light source representing food; (3) individual food- (i.e., light-) seeking performance was evaluated via a fitness function; (4) a genetic algorithm was used to obtain the gene pool of the next generation of robots, with the offspring of the current population inheriting the genetic traits of the fittest parents; (5) the researchers built the next generation of robots, whose notochord stiffness matched the genetic profiles calculated at the previous step; the process was then iterated from Step 2.

In sum Long and his team built generations of robots that differed from one another in terms of the stiffness of their notochords. Stiffness was allowed to evolve from generation to generation by defining a fitness function that rewarded feeding ability. Long is careful to justify why he believes that the evolution of notochord stiffness in his robots may be brought to bear on the evolution of vertebral columns in fishes. Other important methodological aspects whose detailed discussion is beyond the scope of this review—mainly concerning the characteristics of the robot, the genetic algorithm, the fitness function, and the experimental environment—are also carefully discussed and justified in the first part of the book (especially in Chapters 3 and 4).

Note that the methodology implemented by Long is extremely time-consuming. In his study, something like 120 experimental trials were performed, spanning ten generations of robots. The same robots were used in each successive population, but the notochords were replaced from generation to generation, in line with the results of the genetic algorithm. Therefore, in addition to the research effort required to design and build light-seeking robots with the ability to swim, a huge amount of work was needed to build the 360 biomimetic backbones (of variable stiffness) used in the experiments. These efforts were rewarded by a unique opportunity to observe a speeded-up evolutionary process, whose outcome is practically impossible to predict.

On the one hand, in such a scenario it is reasonable to expect that notochord stiffness will change from generation to generation, given that it crucially depends on the performance of each population as calculated by the fitness function. And the performance of each population depends on a number of factors, that includes the stiffness of the fitted notochords but also other environmental factors, which vary from individual to individual and from trial to trial. Over and above this generic prediction of change, however, it is practically impossible to predict how stiffness will evolve, for the very reason that individual performances are unpredictable without computational tools. Will robotic backbones become progressively stiffer in successive generations, thus corroborating the author's hypothesis that vertebral columns evolved from notochords under selection pressure for enhanced feeding? Or will they violate this expectation? The fact that machines—whether robots or computer systems—can behave in unpredictable ways although their sensory-motor control system is algorithmic has important scientific and ethical consequences. Machines are man-made systems, but in order to understand what they are able to do, it is vital to run experiments: their capabilities must be discovered. This point has often been emphasized in the philosophical and computer science literature (see Herbert Simon's *The Sciences of the Artificial* and Valentino Braitenberg's *Vehicles*). Long provides an eminently clear illustration of it in this book.

And what about the second side to the key question about evolution identified earlier—why it is that different forms did not evolve? This question may only be partially answered by making a single population of robots evolve using the method just outlined. A sole instance of the artificial evolutionary process starts from a population characterized by a particular distribution of notochord stiffness and ends (in Long's experiments, at the tenth generation) with a population characterized by a different distribution of stiffness. Thus, from a single instance of implementing this methodology, one can at best learn how notochord stiffness might evolve under a given selection pressure starting from given initial conditions. Could a totally different level of stiffness have evolved starting from totally different initial conditions? A systematic exploration of the evolutionary paths originating from different initial conditions is prohibitive, considering the time and effort required to explore just one path. In Chapter 7, Long explains how so-called "Evolutionary Trekkers" (ETs) can contribute to overcoming this difficulty. Suppose, for example, that one wishes to understand why a particular species of fishes did not evolve backbones with 50 vertebrae. Long's method of investigating this is to build a robotic fish with 50 vertebrae and see, for example, if it survives under a plausible selection pressure. This is an ET, that is to say, a robot representing an individual placed at a specific position on the map of all the possible outcomes of evolution. Systematic failures of a 50-vertebrae ET to find food may lead one to draw the conclusion that a 50-vertebrae backbone cannot guarantee efficient feeding—and that, plausibly, fishes with 50-vertebrae backbones would not have survived under a feeding selection pressure.

Thanks to its detailed treatment of these methodologies, Long's book provides the reader with a clear understanding of how robots are used in evolutionary biorobotics. In addition, he presents arguments justifying the use of robots for this particular purpose, also explaining why he believes that they make a richer contribution to the study of evolution than more traditional experimental methodologies (Chapters 1 and 7). His argumentative strategy is to consider and refute possible objections to evolutionary biorobotics. A first, very general, objection discussed by Long states that direct experimentation on the target living system is preferable over experimentation on a model (i.e., an abstract and idealized representation) of it. To understand the relationship between the speed of fishes and some particular physical parameters of their backbones one may build a swimming robot, intervene on the characteristics of its backbone, and measure the resulting speed in search of a correlation. However, according to the no-model criticism, such an approach can at best result in a highly abstract and idealized view of what actually happens in living fishes. Direct measurement of the parameters of interest are preferable to the use of models if we are to develop good theories of living system behaviors.

Long counters this no-model objection in a very effective and enlightening way. He notes that a fish backbone connected to a measurement apparatus will probably behave differently to the backbone of a fish freely swimming in its ecological niche. Measurements are interventions that unavoidably distort the phenomena of interest, rather than providing a neutral and faithful image of them. More crucially, measurement data provide a view of the system under particular (artificial) conditions that are often very different to the conditions under which the target system ordinarily lives: possible perturbing conditions are controlled or neutralized, and for this reason, they may be abstracted away from the theory being developed. Philosophers of science have often emphasized the non-neutral character of measurements and observation, proposing concepts of scientific theory (notably, the so-called semantic view of scientific theories) that effectively equate theories with models. These considerations clearly undermine the strength of the no-model objection.

After dismissing the no-model objection, Long notes that robotic models may be particularly valuable in the study of evolution given that they—under a number of auxiliary assumptions—enable one to obtain the behavior that extinct species might have produced under particular conditions. However—this is another objection to evolutionary robotics considered by Long—why not use *computer*-based, rather than robotic, simulations? Experiments with computer simulations are much faster than experiments with robots (to change the stiffness of a simulated notochord one only needs to alter a parameter in the computer program). This implies that it is possible to perform a greater number of experiments. Instead of using ETs, one might well explore a substantial portion of the evolutionary map with the aim of assessing whether and why certain forms did not evolve. Computer simulations are also cheaper than robotic simulations. So, why should one choose the evolutionary biorobotics approach? Long describes why his team eventually decided to address particular scientific questions using computer simulations, and acknowledges the aforementioned benefits of computer-based over robotic models. Nonetheless, he is still a firm advocate of the use of robotic systems for the simulation of evolutionary processes. I feel that his arguments in favor of this position are compelling, but not conclusive.

A first reason put forward by Long is that, in order to build a computer simulation of a living system, one must also simulate the environment in which the target system lives. And it is possible to build a physically inaccurate

computer simulation of the environment—or at least, there is nothing to prevent one from building a simulation that violates the laws of physics. On the contrary, the environment of a robotic simulation cannot violate the laws of physics, for the simple reason that it is the real-world environment. This, Long argues, is one of the main reasons for using robotic and not computer-based simulations. I find this argument to be convincing only in part. It is true that computer simulations can violate the laws of physics. However, as acknowledged by Long, they do not need to.

This is not only because today's computers have enough computational power to run physically accurate simulations of realistic environments, but also because some aspects of the physical world may simply be irrelevant to the hypothesis under scrutiny. Recall that robotic models, in (evolutionary) biorobotics, are used to test scientific hypotheses. As far as the study of animal behavior is concerned, scientific hypotheses postulate the existence of particular relationships between particular aspects of the agent's environment and particular (behavioral or evolutionary) phenomena. In fact, scientific hypotheses select and describe those features of the environment that are supposedly causally relevant to—that is to say, that supposedly make the difference with respect to—the occurrence of the target phenomenon. They mention these and only these environmental features. Let us then represent scientific hypotheses as implications of the form $e_1, ..., e_n \rightarrow P$, where $e_1, ..., e_n$ represent the causally relevant environmental factors and P the phenomenon under investigation. Long's passionate support for robotic models may seem somewhat less justified if we note that n may be low in some cases, and that the set of supposedly relevant environmental factors may be relatively easy to simulate on a computer. In such cases, one is in the position of being able to build an accurate simulation of all the physical factors mentioned in the hypothesis—for the simple reason that the environmental factors that are presumed to impact on the phenomenon to be explained are few and easy to simulate.

Consider also that, to evaluate a hypothesis of the form $e_1, ..., e_n \rightarrow P$, one must perform experiments in which only $e_1, ..., e_n$ hold, to see whether P occurs or not (plus other control experiments in which some of the antecedent conditions are varied). Computer simulations offer a unique opportunity to build *ad hoc* environments, where "ad hoc" does not mean that the environment is created in such a way as to obtain data which corroborate the target hypothesis, but rather that the experimental (simulated) environment exactly matches the antecedents of it. In a computer simulation, one may exclude the occurrence of perturbing factors or the intervention of causally relevant environmental factors that are not

mentioned in the hypothesis, and thus include only the supposedly relevant factors. If the simulation accurately reproduces these (i.e., if it simulates accurately enough all the $e_1, ..., e_n$ factors and the relevant aspects of the physical structure of the system), then one may be relatively sure that the behavior of the simulation will be diagnostic of the plausibility of the target hypothesis. The same is not true of robots: they are put "directly" into a non-simulated environment in which there is no guarantee that only $e_1, ..., e_n$ will hold. Long, commenting on roboticist Rodney Brooks, rightly points out that the world is not a model of itself; the world is just the world. In other parts of the book, Long stresses that models are pervasive in—one might well say, constitutive of—scientific research. Hypotheses are typically tested in science by observing the behavior of models of the target system in suitably constrained environments, that is to say, in *models of the environment* of the target system in which the non-relevant aspects of the environment have been neutralized. Thus, the fact that robots are put "directly" into the world, rather than into a model of the world, is not a strong reason for preferring robotic over computer-based simulations.

In Chapter 7, Long adds that while robotic and computer simulations are on a par with respect to being a representation of the target system, the "conceptual distance" between model and target is higher in computer simulations, because they represent both the environment and the behavior of the system while a robot's behavior "just happens". Long points out that a robot is hardware that is situated in the world, while a computer-simulated agent is represented on a piece of hardware (i.e., on a computer, which admits of physical description) situated in the world. For this reason, the "conceptual distance" between the hardware and the target system is higher in the case of the computer simulation than in the case of the robotic model: only in the first case does something else (i.e., a simulated agent) come between the hardware and the target system. However, this argument is not totally convincing, possibly because Long does not explain exactly what he means by "conceptual distance". What he is trying to argue is that robotic simulations are preferable to simulations running on a computer. The fact that robots "are" hardware, while computer simulations have some hardware "behind" them, does not increase the computational distance—whatever that may be—between the simulation and the target. What matters is the distance between the model—that is to say, between that whatever plays the representational role—and the target system, irrespective of whether there are supplementary levels of analysis behind the model.

To summarize, Long's book provides an extremely useful understanding of how robotic systems may be used

to study evolution. More importantly, a large part of it is devoted to justifying the plausibility and the value of these methodologies. While some of Long's arguments may not dispel all doubts on the relative advantages of robotic models versus computer simulations, I feel that the book is to be praised, because it raises a number of urgent methodological and epistemological issues that apply both to evolutionary biorobotics and to biorobotics at large.

As earlier flagged, a second reason that makes this book particularly interesting, especially for readers with no direct experience of scientific research, is that Long devotes considerable space to the discussion of what may be termed a case of falsification: experiments with evolving robots produced results that were at odds with the predictions made by the research team. Long's analysis of this case, in my opinion, brings to the fore some important facts about science in general.

To illustrate, let us recall that although it was reasonable to predict that the Tadro3s' notochord stiffness would change during the evolutionary process, it was impossible to forecast the direction of that change. However, Long and his team had good reasons to believe that selection for enhanced feeding behavior would make stiffer notochords evolve from generation to generation. To their disappointment, however, they realized at the end of the cycle of experiments that this had not been the case (Chapter 4). Feeding efficiency decreased instead of increasing in some cases, even though it had been rewarded via the fitness function. And more crucially, changes in feeding efficiency were unrelated in some cases to changes in the stiffness of the notochord. There were even cases in which feeding efficiency increased while stiffness decreased, in striking contrast with the researcher's initial expectations. Situations of this kind are often called cases of *falsification*.

"Strictly speaking", Long claims, "you can only falsify a hypothesis—you can't prove it" (p. 161). By this statement, he presumably means to say that one is never in a position to conclusively assert that a hypothesis is true, while one may be in a position to conclusively assert that a hypothesis is false. However, when interpreted in this way, the author's claim is false: there is no such asymmetry between hypothesis corroboration and falsification. It is true that one cannot conclusively prove a hypothesis, given that the possibility of future counterexamples cannot be rationally excluded. However, neither can it be rationally concluded that a hypothesis is false. Indeed, "bad" experimental results may be due to factors unrelated to the hypothesis being tested—for example, inaccurate experimental apparatus—and thus not indicative of the falsity of the hypothesis. Researchers can strive to exclude this possibility (for example, by double-checking

the accuracy of their experimental apparatus) but it can never be conclusively discounted, given that checking procedures may themselves be inaccurate, or additional issues may have been overlooked. A hypothesis may only be considered falsified based on the best available evidence. There is a clear symmetry with corroboration in this respect, as hypotheses can also only be accepted based on the best evidence currently available. The only actual asymmetry between falsification and corroboration is emotional: when researchers obtain data that apparently contradicts a hypothesis that they had formulated, they typically experience disappointment. The same is not true of corroboration.

In the case in hand, as discussed in Chapter 4, Long and his team concluded that their unexpected results were due to having chosen a mistaken fitness function (see the third step in the evolutionary biorobotics method described above). The function used in the experiments rewarded the *speed* of the robot, the *time* it took to reach the light (a higher score was awarded for a shorter time), the *distance* of each robot from the light at the end of the trial (more points were given for shorter distances) and the amount of body wobble (smaller amounts of wobble were rewarded, under the assumption that wobbling is energy-wasting). This function mathematically expressed the researchers' initial view as to what constitutes efficient feeding behavior. However, after careful methodological reflection and additional experiments, they eventually realized that this view was wrong. In particular, they obtained evidence suggesting that body wobble may favor feeding: wobbling does not only require a lot of energy, it also enables the system to make rapid turning maneuvers that can help to find food (and avoid predators). Long explains why he believes that, had body wobble been rewarded and not penalized by the fitness function, the experimental results would not have been so inconsistent with the initial hypothesis.

Long's discussion of this case illustrates the fact that the numbers produced by the measurement tools used in scientific research do not provide theoretically neutral and unequivocal descriptions of reality. As commonly stressed by philosophers of science, experimental data are often massively theory-laden. In the case just outlined, the numbers produced by the fitness function did not describe the Tadro3s' performance in a neutral and unequivocal way: they depended on an implicit view (i.e., on an implicit theory) of what counts as an efficient feeding behavior. Change the background theory, and the data on the same robotic behaviors will change. Professional scientists are well aware of the theory-laden character of scientific data, but people who are not directly involved in scientific research may still hold the popular and mythical view of experimental data as neutral

representations of reality. Long's presentation of this case should therefore help this class of potential readers to appreciate what scientists are really talking about when they refer to "data".

Long's book may also be insightful for this class of readers in that it illustrates the *recursive* nature of scientific research. Scientists conduct experiments to test a theory. Then, they must test the theories on which the experimental results rest (which in Long's case included his initial view of what constitutes an efficient feeding behavior) in order to interpret them correctly. To test these background theories, they often need to run additional experiments—this was the case in Long's study, given that his revised view of the function of body wobbling resulted from a separate and successive class of experiments testing the Tadro3s' behavior. Such additional experiments, in turn, rest on other background theories. In practice, one must halt this recursive process at some point, otherwise no result would be ever produced. In principle, however, as illustrated by Long's discussion, science is a never-ending enterprise. Perhaps a future evolutionary bioroboticist will find reasons to reject the background assumptions on which Long's results rest, at some level of the hierarchy. However, we can be sure of one thing: this rejection too will rest on further fallible background assumptions.

Finally, I would like to point out that Long's detailed treatment of this particular case of data that failed to satisfy predictions illustrates how experimental failures can occasionally lead to interesting discoveries: while trying to explain the incongruent results obtained, the researchers ultimately discovered something new about the function of body wobbling in fish. Unexpected and disappointing results can potentially yield unexpected and gratifying progress. In this review, I have argued that Long's book provides a valuable introduction to the field of evolutionary robotics. In addition, as drawn out in these concluding remarks, I believe that it also offers an insightful discussion of the methodological complexities ordinarily faced by scientists in the course of their discovery processes, illuminating key aspects of scientific research at large. These are both excellent reasons for reading it.