Roger Penrose · Emanuele Severino ·
Fabio Scardigli · Ines Testoni · Giuseppe Vitiello ·
Giacomo Mauro D'Ariano · Federico Faggin *Authors*
Fabio Scardigli *Editor*

Artificial Intelligence Versus Natural Intelligence

Springer

Artificial Intelligence Versus Natural Intelligence

Roger Penrose ·
Emanuele Severino ·
Fabio Scardigli ·
Ines Testoni ·
Giuseppe Vitiello ·
Giacomo Mauro D'Ariano ·
Federico Faggin
Authors

Fabio Scardigli
Editor

Artificial Intelligence Versus Natural Intelligence

Springer

Authors
Roger Penrose
Mathematical Institute
Oxford University
Oxford, UK

Fabio Scardigli
Dipartimento di Matematica
Politecnico di Milano
Milan, Italy

Giuseppe Vitiello
Dipartimento di Fisica
Università di Salerno
Fisciano, Italy

Federico Faggin
Los Altos Hills, CA, USA

Emanuele Severino
Brescia, Italy

Ines Testoni
FISPPA, Section Applied Psychology
University of Padua
Padova, Italy

Giacomo Mauro D'Ariano
Dipartimento di Fisica
Università di Pavia
Pavia, Italy

Editor
Fabio Scardigli
Dipartimento di Matematica
Politecnico di Milano
Milan, Italy

ISBN 978-3-030-85479-9 ISBN 978-3-030-85480-5 (eBook)
https://doi.org/10.1007/978-3-030-85480-5

Language rights for Italian edition excluded.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Introduction

Fabio Scardigli

This book contains the transcriptions of the talks and the debate between Roger Penrose and Emanuele Severino that took place during the conference "Artificial Intelligence vs Natural Intelligence", held in Milano, at the Cariplo Congress Center, on May 12, 2018.

Besides the keynote speeches of Penrose and Severino, there were the illuminating talks of Giuseppe Vitiello (theoretical physicist), Mauro D'Ariano (theoretical physicist), and Ines Testoni (psychologist), which gave rise to the three essays completing this book.

The conference was conceived and organized (like the previous one on "Determinism and Free Will") by a group of friends and colleagues: Fabio Scardigli, Marcello

F. Scardigli (✉)
Department of Mathematics, Polytechnic of Milan, Milan, Italy
e-mail: fabio@phys.ntu.edu.tw

R. Penrose et al., *Artificial Intelligence Versus Natural Intelligence*,
https://doi.org/10.1007/978-3-030-85480-5_1

Esposito, and Marco Dotti. Our warmest thanks go to our colleague Massimo Blasone for his help during the workshop's days.

The success of the conference was somehow astonishing, even greater than that of the previous meeting. More than 600 people crowded into the main hall of the Cariplo Congress Center and into two adjacent rooms, equipped with closed circuit television. This vividly testifies to the great interest that the general public has for the themes of Artificial Intelligence, Theory of Consciousness, Intelligent devices, and all that.

1 Understanding and Algorithms

Regarding those topics usually grouped under the heading "AI" (Artificial Intelligence), the perspectives of the two main speakers, the mathematical physicist Roger Penrose and the philosopher Emanuele Severino, are obviously quite different. Nevertheless, as the reader will soon discover, both agree that we do not yet have "intelligent devices", and also that, if we follow the vision of the so-called Strong AI (presently still the mainstream), we will never be able to build such devices. This opinion is also supported by the authors of the other essays in the book, although with their own slant.

In his main talk, Penrose focuses on the relations between the words "intelligence", "understanding", and "consciousness". "Being a mathematician", he kids, "connections among words are more important to me than their "true" meaning". Relations among concepts are more illuminating than substantial definitions, in other words. Therefore Penrose starts from the idea that the word "intelligence", at least in the standard usage, implies

"understanding", and "understanding" requires some "awareness" or "consciousness". Going through different examples, discussed in great detail, Penrose shows that the machines, or software, at our disposal today are computational devices, maybe very sophisticated, but all essentially based on the ideal prototype of the Turing Machine. "Intelligence" and "understanding", on the other hand, seem to exhibit properties that escape simple computability. Intelligence is something more than mere computational ability. Examples taken from chess, mathematical induction (the intriguing Goodstein theorem), the tiling of the Euclidean plane (polyominoes, which cannot be produced through computable algorithms), all converge to show that "understanding is something which is not achieved by rules". So, a general quality of understanding seems to be that it is not an algorithm. Hence, "understanding", according to Penrose, is not an element or a result of a (very) complicated application of a set of rules (algorithm).

2 Quantum Mechanics and Consciousness

Penrose then reviews the two main theoretical pillars of modern physics, namely General Relativity and Quantum Mechanics, and points out that there is only one specific element of modern physics that cannot be reproduced on a computer, because it is not computable. To be precise, it is the measurement process in Quantum Mechanics, the so called "collapse of the wave function". This is not described by the Schroedinger equation, and cannot be implemented (not even in principle) by any computable algorithm.

In the words of Penrose "The idea is that the collapse process is something which is not computable. In fact it is something which is a bit like what free will might be. Because according to the current physics, it is making its own decisions. Somehow it decides to be here or there."

Thus, in the whole landscape of physical theories and phenomena, there seem to be only two things characterized by their intrinsic non-algorithmic, non-computable nature: the measurement process in quantum mechanics (or the collapse of the wave function), and the phenomenon of "understanding" or "awareness" proper to (human) consciousness. In his work, Penrose has been trying to connect these two concepts from many years now. Starting in particular with the book "The Emperor's New Mind", he formulates a theory according to which the origin of the non-algorithmic processes of "understanding", "awareness", and "consciousness" should be sought in certain quantum processes occurring in specific regions of the brain. Thanks to his collaboration with the anesthesiologist Stuart Hameroff, during the 1990s, some promising candidates to host these so-called "Spontaneous Orchestrated Reductions" of the quantum states were identified in the "microtubules". Microtubules are very small sub-cellular structures, located inside neurons, in the axons and dendrites.

In these structures, quantum coherence could be maintained long enough to allow the quantum collapse of a wave function to directly influence "elements", or "atoms", of quantum consciousness, named "protoconsciousness". According to Hameroff and Penrose, these could be the building blocks out of which consciousness is constructed. Obviously, in these building blocks, there is not yet a conscious aim, not yet a purpose, and even less a meaning, but out of them structures containing conscious behavior might emerge.

3 Protoconsciousness

An immediate consequence of this theory is that consciousness, far from being a purely human characteristic, should instead emerge wherever structures like microtubules are present, hence, for example, in animals like apes, dolphins, dogs, cats, or mice. This vision, also full of ethical implications, was what sparked the discussion with Severino. Moreover, Penrose also briefly introduced the possibility of "constructing" a conscious, and therefore intelligent, device. In his perspective, of course, this could be done only by aggregating elements of protoconsciousness and providing them with the appropriate environment for an "orchestrated objective reduction" of the wave function to take place.

Quite obviously, according to Penrose, insofar as they are based on purely algorithmic capabilities, the computers of today, and most likely also those of tomorrow, are and will remain "unconscious", hence lacking in real understanding and intelligence. In particular, there is no danger that "intelligent machines" will one day be able to take over the world, and threaten or destroy the human race. On this issue, it is well known that Stephen Hawking used to have an opposite view.

4 Free Will and Singularities

In connection with the theme of free will, and in the light of the above ideas, we can attempt some further consideration. As it is well known, professional general relativists are afraid of singularities, which can appear in General Relativity. Roger Penrose won the 2020 Nobel prize in Physics just because in his seminal paper of 1965 proved a

theorem which stated that, under very general conditions (in particular, without requiring any spherical symmetry), singularities are a general and inevitable prediction of General Relativity (both in the past direction, i.e. the Big Bang cosmological singularity, and in the future direction, i.e. the black holes singularities).

But why theoretical physicists don't like singularities?

The standard answer is that in a singularity the predictive ability of the theory fails completely. From a singularity literally anything could come out (or enter), in a completely unpredictable way, since physical laws crash down, by definition, in a singularity. In fact, to protect the observable universe from such a monster, Penrose proposed long ago the cosmic censorship conjecture, according to which singularities are always wrapped and hidden behind an event horizon. And therefore they cannot influence the external universe.

However, according to the above considerations of Penrose, there is also another phenomenon that presents points of "singularity", where the predictive capacity of the theory fails completely: it is the single event in quantum mechanics! We have crashes of computability at every collapse of the wave function, therefore "singularities" where the predictive ability of quantum theory fails. If for example we consider the emission of a single photon from an excited atom, well, it is not possible to predict either the instant or the direction of such emission. And analogously happens for the decay of a neutron: neither the instant nor the direction of emission of the pair neutrino-electron are known, for a single event. Quantum theory predicts only probability distributions, which, although in perfect agreement with experiment, by definition, apply to classes of events only, not to single events. Therefore, "singularities", namely points where predictability crashes, appear to be present everywhere in Quantum Theory.

5 Constructing Consciousness

Two of the themes which Emanuele Severino picks out for criticism are the "place" in which human intelligence or consciousness might be located, and the possibility of "constructing" an intelligent (or conscious) device.

Severino starts by pointing out the hypothetical-deductive character of Science. All sciences, and in particular the mathematical sciences, are based on postulates, that is, propositions that are taken for granted, from which, according to certain (logical) rules, other propositions will follow. But postulates and even deductive rules are not considered to be incontrovertible truths, even by Science itself. In particular, they are themselves "conventions". Or, to use a word to which Severino attributes a wide meaning, they are "faiths". That is, expressions of the "will to power". According to Severino "the choice between two different and competing theories is ultimately determined by the ability of one rather than the other to transform the world". Science itself recognizes that it is no longer possible to build a knowledge that no one, either gods or men, could deny. Therefore, in Severino's view, Science does not aim to find incontrovertible truth; it aims rather to have power over the world.

This objective is opposed to what Severino maintains to be the primeval, foundational goal of philosophy, the goal for which philosophy was born 2500 years ago in Greece, namely to "unveil" incontrovertible truths.

6 Mathematical Modeling

However, Severino's view that the ancient Greek mathematicians, and others right up to Galileo, were trying to arrive at incontrovertible epistemic truths ("to know theorems as

God knows them", in Galileo's words), can surely be criticized. It is in fact rather clear, especially according to recent historical reconstructions (see, e.g., Lucio Russo), that the concept of "mathematical model" was already well developed in the work of Euclid, Archimedes, and others. The free choice of postulates, the possibility of "playing around" with them in order to better model a given phenomenon, or a given set of (mathematical) propositions, are all operations very clearly stated and effectively performed by ancient mathematicians, right down to the modern fathers of the scientific revolution, Copernicus, Galileo, Newton, etc. Actually, it can be safely affirmed that not one of them really believed in the construction of absolute truths; instead they were concerned with (mathematical) models performing better than the previous ones (e.g., finding a physics better than Aristotelian physics, or an astronomy better than Ptolemaic astronomy, etc.) (See also Determinism and Free Will, Introduction, on this point).

7 Manifestation of the World

Severino then introduces the concept of "manifestation of the world", the dimension where everything happens, the dimension where everything manifests itself, the dimension from which the news of every single fact or thing is drawn: "There is no step that science can take that does not spring from the manifestation of the world." According to Severino, the "manifestation of the world" is not a thing among other things, because it contains all the past, present, and future things of the world, everything that appears.

Severino appeals also to the idea of "manifestation of the world" in his second intervention, where he strongly opposes Penrose's view of "a place (in the brain) where to

look for consciousness". According to Severino, looking for a place "where consciousness resides" means, in fact, considering consciousness once again merely as a specific part of that "appearing of the world", of that "manifestation of the world", which should actually be considered as the primary form of consciousness.

Nevertheless, this idea of a "consciousness of the world" would appear to share traits with the "atoms" of proto-consciousness (common to different entities) introduced by Penrose himself, and similarities can also be found in concepts considered by the other authors in this book, Vitiello, D'Ariano, and Faggin, although with different emphases and viewed from different angles.

8 Production

The core of Severino's criticism of the idea of "artificial" intelligence, or "artificial" consciousness, focusses on the idea of "production" itself. As is customary in Severino's philosophical position, the error, the major nihilistic mistake of Western philosophy (civilization), hides within the verb "produce", in the idea of "construction", i.e., of the "production" of something. Far from being innocent, as it may appear, the concept of production, "poiesis" in Greek, hides the obvious, common-sense belief that things can be made to come out from nothing, and (if desired) to go back, to return to nothing. According to Severino, the observation that things can be created and destroyed, the belief that things oscillate between being and not being, with the associated transition from not being to being, and therefore the becoming of things, which we all consider to be the supreme obviousness, is not an observable subject, it is not the substance of an observation, but is a theory, namely only one possible interpretation of reality, among

many. Behind this stubborn belief, the ghost of nihilism hides itself: the belief that things are actually nothing.

Clearly, from this point of view, the idea of producing an artificial intelligence is just another manifestation of the deep nihilistic vision pervading Western civilization.

9 Artificial Intelligence

But Severino goes further. He observes that the Platonic definition of "production" as "the cause that makes a thing to go from not being to being" permeates the whole of Western thought, in philosophy, economics, law, and the mathematical sciences, and is always considered as obvious. Moreover, Western thought views Man as the only being able to organize means for the production of purposes. But, as Severino notes, this is also the definition of the machine! Except that, for the moment, machines have no purposes in their sights: "Man is that machine which organizes means in view of the production of purposes, having in mind precisely the presence of purposes, the ideal presence". In this way Severino reaches his provocative conclusion that the "natural" Man is thought of as a machine, and indeed that the world itself is thought of as a mechanism in which means are organized in view of the production of goals. And therefore, given the way in which Man has been understood in the West, then Man, or rather the natural intelligence of Man, "is" already an "artificial" intelligence.

In opposition to that, Severino reaffirms that "the manifestation of the world as a whole—this primary and fundamental form of consciousness—cannot be an object of production, at least for this reason: because the producer, if he had to produce the totality of the manifestation of the world, would have to be outside the

manifestation of the world, and therefore it would be something unknown."

10 Debate

The second part of the dialog contains the interaction, or debate, between Penrose and Severino, and between the public and the speakers. Many questions involved the issue of microtubules. Regarding their capacity to survive death, and hence preserve some memories of the past life, Penrose is definitely skeptical: "I think that microtubules will not survive death any more than neurons". A second set of questions concerned tests of consciousness, and in particular ways to investigate how general anesthetics act on consciousness. The general opinion is that microtubules are very much involved in the actions of general anesthetics. Penrose also introduces the role of the cerebellum here, which controls all the "automatic" actions of a human being, and whose action seem to be entirely unconscious, as opposed to the role of the cerebrum, where conscious actions seem to play an explicit role. As confirmation of his ideas, Penrose points out that microtubules are abundant in pyramidal cells, which are in turn abundant in the cerebrum, while on the contrary they are not found in the cerebellum. What is interesting here is that consciousness seems to appear in pyramidal cells. Other tests of conscious understanding are carried out with specific chess problems, designed for humans and for computers. Here one sees the distinction between "understanding", e.g., what a barrier of pawns does, and the purely "mechanical" calculation which a computer performs. These tests clearly illustrate the difference between conscious thinking (or conscious understanding), and mere computation.

11 Consciousness in Animals

On the contrary, according to Penrose, the notion of creativity is misleading and ambiguous: creativity is not a good test for consciousness. Understanding is something where you can see the difference between conscious or unconscious action. Instead, it is very hard to see the difference between creativity and just random production of something (which is considered different from what has been "created" before). So, creativity, as opposed to understanding, is extremely hard to test, and it is difficult to be objective about that. Finally, Penrose supports the idea that the phenomenon of consciousness is present also in animals. This is consistent with his vision of microtubules as the arena in which protoconscious events, and finally consciousness itself, draw their origin. Microtubules are very present in many "superior" animals, and so therefore should be consciousness, in particular in dogs, cows, elephants, monkeys, gorillas, dolphins, mice, etc. From this point of view, ethical conclusions could be drawn, such as the respect we owe, not just to other humans, but also to (many) other creatures.

12 Consciousness and Language

The idea proposed by Penrose and Hameroff (consciousness emerges out of a cumulative process of elementary "protoconsciousness" acts, until we finally arrive at the wonders of human mind), calls for a disturbing observation about the "strong" AI program. As is well known, and as has happened historically, such a program starts from (formal) languages and aims to rebuild "intelligence" through software (i.e., computer programs based

on languages), following a kind of top-down process. And this is done with the fairly overt conviction that perhaps from the construction of intelligence we can then pass, by continuing on the same path, to the construction of consciousness.

However, according to the ideas put forward by Penrose, Nature seems to arrive first at the construction of simple forms of consciousness, which actually appears to be a fairly common phenomenon at least in the "higher" animals. Only after that, Nature builds languages (even complicated ones); and languages, at least the advanced ones, seem to be relevant only for a single species, humans.

From this point of view, the Strong AI program looks truly "artificial", in the sense that it is moving along a path opposite to the one followed by natural evolution. Humans are trying to build consciousness starting from language, whereas Nature has built languages starting from (proto)consciousness!

On the other hand, as D. Hofstadter once said, perhaps Artificial Intelligence should be compared to a modern jet airplane: high-performance when it comes to specific tasks, but generally unable to do things that a sparrow, which in the metaphor represents human intelligence, can easily do. A jet can fly from London to Milano in a hour, an impossible task for a sparrow or a pigeon. But try to land a jet on a gutter…

13 A Place for Consciousness?

Severino's counter reply to Penrose well illustrates the distance and the differences between their approaches. Severino openly criticizes Penrose for ignoring his words about the "manifestation of the world as a primary form of consciousness", although this concept is present, according

to Severino, in many important exponents of contemporary culture, such as Descartes, Kant, Brouwer. When Penrose repeatedly looks for the "place where consciousness resides", he simply shows, in Severino's view, that he still imagines consciousness as a thing among things, namely only as a part within that "manifestation of the world" which is actually the primary form of consciousness.

A final topic where the disagreement is particularly clear is the practical nature of science, which according to Severino means this: the conceptual articulations of scientific knowledge allow a power over the world superior to other conceptual articulations, such as the conceptual articulation of the alliance with the sacred, i.e., prayer. So, the conceptual articulation of modern scientific knowledge is formidable! It is what today allows the greatest power over the world. But power is one thing, truth is another. Severino comes back to his initial argument about the technological power of scientific theories. His conviction is that a scientific theory ultimately is evaluated on the basis of its technical capacity to transform the world, not its ability to truly represent it, or effectively explain it. The theme of intersubjectivity in science is recalled here by Severino. According to Popper, for there to be power, intersubjective recognition of the power in question is needed, and this means that others have to clearly perceive the transformation of the world.

14 Science and Technology

On this last point Penrose replies forcefully (and in my opinion correctly), by reaffirming the traditional separation between science and technology. Deep ethical issues enter the discussion at this point. In Penrose's words: "My way of looking at science is to try find out what is true

about the world. So, there is no moral issue involved. I mean, the moral issue is a separate question." Science tries to understand the way the world operates, how it works. Then, there is technology, which works on how to use scientific knowledge. Technology has a close relationship with science, but it is not science. Technology has a huge and continued impact on what people do in their everyday lives. And this of course raises the moral issues. So, according to Penrose, the use of technology, and in particular the good or bad use of science, is deeply entangled with morality. Penrose is clear about that: "When I'm doing science, I'm trying to develop an understanding of the ways world operates, and I am not looking for power". Science is not trying to control the world. That is the aim of technology. Technology and morality are areas clearly separated from science, although they depend on science; when science changes, these two other areas have to pay respect to science, and see what it tells them.

Finally, Penrose points out once again that his and Hameroff's ideas would tend to consider consciousness as a phenomenon that is not restricted to human beings, but appears also in other animals. This implies that a moral issue is involved in how we deal with animals, and this is a further example of the way science influences our moral beliefs.

15 Chatbots

The essay by Ines Testoni, a former disciple of Severino, presents a number of interesting points, of which I can only discuss a few here. The essay opens by presenting an experiment, actually performed in 2017 at the Facebook Artificial Intelligence Research (FAIR) group. Two computer programs (software) were trained in English conversations, and then allowed to chat autonomously with each