PENULTIMATE DRAFT – PLEASE CITE THE PUBLISHED VERSION

To appear in: *A Critical Reflection on Automated Science*, Bertolaso, M., Sterpetti, F. (eds.), Springer.

Introduction. Human Perspectives on the Quest for Knowledge

Marta Bertolaso

Campus Bio-Medico University of Rome

Faculty of Engineering & Institute of Philosophy of Scientific and Technological Practice

m.bertolaso@unicampus.it

Fabio Sterpetti

Sapienza University of Rome

Department of Philosophy

fabio.sterpetti@uniroma1.it

Abstract

We firstly introduce the new Springer series *Human Perspectives in Health Sciences and Technology* (HPHST), and then we move on to illustrate the topic *this volume* deals with, namely whether machines will replace scientists in scientific development. We then explain the decision of having *this volume* to be the first volume of the HPHST series. Finally, we describe the organization of *this book* and give a brief presentation of each chapter.

Keywords

Automated Science; Epistemology; Health Sciences; Scientific Discovery; Scientific Practice; Technology

1. Introducing the New Series

This volume is the very first one to appear in the new Springer series *Human Perspectives in Health Sciences and Technology* (HPHST). It is first of all worth clarifying that the term 'human' in the series' title has not to be understood as a synonymous of the term 'humanities' as it is usually understood in the extant literature. Indeed, the HPHST series is not another series devoted to what is usually referred to by the expression 'medical humanities', which nowadays is a quite precisely delimited disciplinary field.

To appear in: *A Critical Reflection on Automated Science*, Bertolaso, M., Sterpetti, F. (eds.), Springer.

The HPHST series aims to provide an editorial forum to present both scientists' cutting-edge proposals in biomedical sciences which are able to deeply impact our human biological, emotional and social lives, and thought-provoking reflections by scientists and philosophers alike on how those scientific achievements affect not only our lives, but also the way we understand and conceptualize how we produce knowledge and advance science, so contributing to refine the image of ourselves as human knowing subjects. The main idea that led to the creation of the series is indeed that those are two sides of the same 'being-human' coin: scientific achievements can affect both our lives and ways of thinking, and, on the other hand, a critical scrutiny of those achievements may suggest new directions to scientific inquiry. So, although epistemological, social, and ethical issues are certainly all central for the series, what distinguishes it from other already existing series dealing with similar issues, is that HPHST aims to address those issues from a less rigidly pre-defined disciplinary perspective and be especially attractive for non-mainstream views and innovative ideas in the biomedical as well as in the philosophical field.

The HPHST series aims to deal both with general and theoretical issues that spread across disciplines, such as the one we deal with in *this volume*, and with more specific issues related to scientific practice in specific domains. The series focuses especially, although not exclusively, on health sciences and technological disciplines, since technology, bio-medicine, and health sciences more in general are coevolving in unprecedented ways, and much philosophical work needs to be done to understand the implications of this process.

Technological development has always offered new opportunities for scientific and social advancement. In the last decades, technology has entered in intimate partnership with the life sciences, providing tools to isolate and modify experimental systems in vitro, offering computational power to expand our cognitive capacities and grasp features of complex living systems, then introducing digital models and simulations, providing methods to modify and 'rewrite' certain life processes and organismal traits, and allowing more daring and smooth hybridizations between artificial design and natural systems.

Incredible improvements to human life have come from these techno-scientific developments. Complex ethical questions have also arisen regarding the impact and the use of these developments. The life sciences have been shaking their paradigm and transforming the ways we conceptualize and deal with organisms and living systems, including humans (see e.g. Soto, Longo, Noble 2016). Bio-medicine is also experiencing ever increasingly difficult challenges (see e.g. Ioannidis 2016). In fact, we live in a time

of widespread worries and fears about science, and also of scepticism and resignation regarding the extreme complexity of living and social systems. Great expectations are placed on technology, and much of science is said to be technology driven, but human choices and responsibility remain ineliminable ingredients in the task of elaborating on the acquired knowledge and thus improving our scientific understanding about the natural world. A driving persuasion of HPHST series is that a new trust in science is possible, but it must be based on a sound and up-to-date epistemology, and a recognition of the inherent ethical dimension of science as a human endeavor: the only way to understand and govern science and communities for the better is a view of science rich in all its human aspects, requiring the contribution of philosophy, as well as natural and social sciences. Our hope is that the HPHST series can contribute to such a renewed trust in science and to a real improvement of it.

2. The Theme of the Volume

This volume originated from the conference "Will Science Remain Human? Frontiers of the Incorporation of Technological Innovations in the Bio-Medical Sciences," which took place in Rome, at the University Campus Bio-Medico, in March 2018, thanks to the support of the Social Trends Institute. When working on *this book*, we decided to deal with the very same challenging question, namely whether science will remain human, but to focus a little bit more sharply, although not exclusively, on the issue of whether science will remain human notwithstanding the increasing automation of science. So, we thought, together with Floor Oosting, Springer Executive Editor of Applied Ethics, Social Sciences, and Humanities, whom we wish to thank for all her support and advice, that *A Critical Reflection on Automated Science – Will Science Remain Human?* would be an appropriate title for *this book*.

According to some philosophers and scientists, humans are becoming more and more dispensable in the pursuing of knowledge, since scientific research can be automated (Sparkes *et al*. 2010; King *et al*. 2009; Anderson 2008; Allen 2001). More precisely, according to some philosophers and scientists, the very aim of Artificial Intelligence today is not merely to mimic human intelligence under every respect, rather it is automated scientific discovery (Sweeney 2017; Colton 2002). Those claims are very appealing and ever more shared by many scientists, philosophers and lay people. Yet, they raise both epistemological and ethical concerns, and rely on assumptions that are

disputable, and indeed have been disputed. From an epistemological point of view, consider, for instance, that assuming that scientific discovery can be automated means to assume that ampliative reasoning can be mechanized, i.e. that it is algorithmic in character. But proving that this is the case is not an easy task (see e.g. Sterpetti, Bertolaso 2018). From an ethical standpoint, consider, for instance, issues of responsibility on data management processes (which entail consistency of data organization and transmission with the original scientific question, contextualization of data, etc.) or on possible (and often unforeseen) consequences of completely automated researches. As an analogy, think of the difficulties we have in acknowledging limits of the majority of target therapies, as well as of the debate on who is responsible for deaths provoked by self-driving cars. So, although there is an increasing enthusiasm for the idea that machines can substitute scientists in crucial aspects of scientific practice, the current explosion of technological tools for scientific research seems to call also for a renewed understanding of the human character of science, which is going to be sometime less central but not for this less fundamental.

The topic *this volume* deals with, namely whether the computational tools we developed in last decades are changing the way we humans do science, is a very hot topic in the philosophy of science (see e.g. Gillies 1996; Humphreys 2004; Nickles 2018). The question that many are trying to answer is the following: Can machines *really* replace scientists in crucial aspects of *scientific advancement*? Despite its interest, it is a topic on which, to the best of our knowledge, one can find very few consistent works in the literature. This is why *this volume* brings together philosophers and scientists with different opinions to address from different perspectives the issue of whether machines can replace scientists. The book's aim is to contribute to the debate with valuable insights and critical suggestions which might be able to further it toward more reasoned and aware stances on the problem. Another feature of *this volume* that might be interesting for readers, is that it does not only deal with the issue of automated science in general, but it also focuses on biology and medicine, which are often ignored when abstract and general issues such as whether scientific research can be automated are discussed.

Most of the times, works that are devoted to the topic at stake try to support or deny the hypothesis that science can be automated from an 'engaged' perspective. On the contrary, *this volume* tries to scrutinize that hypothesis without prejudices or any previous theoretical commitment. There are indeed different opinions among the contributors to *this volume* on the automation of science, but overall the book integrates reasons for thinking that current computational tools are changing our way to do science, while they

also ask for a deeper philosophical reflection about science itself as a rational human endeavor. This opens to innovative thoughts about the peculiar way humans know and understand the world and gives us reasons for thinking that the role of the human knowing subject in the process of scientific discovery is not deniable nor dispensable.

We decided that *this volume* would be the first one of the HPHST series for several reasons. First of all, *this book* shares the focus of the series on the interplay between the human subject and technology. As we said, between the great promises of technology and the great fears that technology may replace the human in driving important aspects of our life, to get the right attitude for future scientific practice we need to develop an adequate epistemology. To this end, the series invokes contributions from philosophy as well as natural and social sciences, and this is exactly what *this book* aims to provide. Also, *this volume* deals with the human subject as an object and a subject of inquiry, an approach that perfectly fits with the series' approach. Indeed, in order to claim that scientific discovery can be automated, scientific discovery needs to be clearly understood. And in order to understand scientific discovery, it is usually believed that human reasoning needs to be clearly understood. This means that we need to improve our understanding of a qualifying feature of human nature, i.e. reasoning, and of what do humans really do when they do science. Moreover, *this volume* deals with epistemological, ethical, and technological issues, and focuses specifically on the bio-medical sciences and technology. Finally, in accordance with the series' aims, *this volume* tries to provide both scientists' and philosophers' reflections on practical and theoretical aspects of science, and so to further our understanding of how we produce knowledge and advance science. For all those reasons, it was natural for us to think about this volume as the most adequate one to inaugurate the HPHST series.

3. Overview of the Volume

The book is divided into three parts. The first part, *Can Discovery Be automated?*, addresses the question of whether scientific discovery can be automated from a general and theoretical perspective. This part consists of five chapters.

The first chapter is Paul Humphreys' *Why Automated Science Should Be Cautiously Welcomed*, which focuses on the notion of 'representational opacity', a notion Humphreys develops in analogy with the notion of 'epistemic opacity' that he introduced in previous works, in order to clarify in what sense the introduction of automated

methods in scientific practice is epistemologically relevant. Humphreys argues for a moderately optimistic view of the role that automated methods can play in the advancement of scientific knowledge. He also draws some interesting parallels between the problem of scientific realism and the problem of internal representations in deep neural nets.

In the second chapter, *Instrumental Perspectivism: Is AI Machine Learning Technology like NMR Spectroscopy?*, Sandra Mitchell addresses the issue of whether something crucial is lost if deep learning algorithms replace scientists in making decisions, by considering whether the ways in which new learning technologies extend beyond human cognitive aspects of science can be treated *instrumentally*, i.e. in analogy with the ways in which telescopes and microscopes extended beyond human sensory perception. To illustrate her proposal, Mitchell compares machine learning technology with nuclear magnetic resonance technology in protein structure prediction.

In chapter three, *How Scientists Are Brought Back into Science – The Error of Empiricism*, Mieke Boon argues that despite machine learning might be very useful for some specific epistemic tasks, such as classification and pattern recognition, for many other epistemic tasks, such as, for instance, searching for analogies that can help to interpret a problem differently, and so to find a solution to that problem, the production of *comprehensible* scientific knowledge is crucial. According to Boon, such kind of knowledge cannot be produced by machines, since machine learning technology is such that it does not provide understanding.

In the fourth chapter, *Information at the Threshold of Interpretation: Science as Human Construction of Sense*, Giuseppe Longo investigates the origin of an ambiguity that led to serious epistemological consequences, namely the ambiguity concerning the use of the concept of 'information' in artificial intelligence and biology. According to Longo, there is still a confusion between the process of knowledge-production and that of information-processing. Science is dehumanized because information is thought to be directly embedded in the world. In order to avoid this shortcoming, we must distinguish between information as formal elaboration of signs, and information as production of meaning.

The fifth chapter is Fabio Sterpetti's *Mathematical Proofs and Scientific Discovery*. Sterpetti claims that the idea that science can be automated is deeply related to the idea that the method of mathematics is the axiomatic method. But, he argues, since the axiomatic view is inadequate as a view of the method of mathematics and we should prefer the analytic view, it cannot really be claimed that science can be automated.

Indeed, if the method of mathematics is the analytic method, then the advancement of mathematical knowledge cannot be mechanized, since non-deductive reasoning plays a crucial role in the analytic method, and non-deductive reasoning cannot be mechanized.

The second part of the book, *Automated Science and Computer Modelling*, deals with an analysis of the consequences of using automated methods that are more focused on biology, medicine and health technologies. In particular, some epistemological issues related to the role that computer modelling, computer simulations and virtual reality play in scientific practice are discussed. This part consists of five chapters.

Fridolin Gross's *The Impact of Formal Reasoning in Computational Biology* investigates the role played by computational methods in molecular biology by focusing on the meaning of the concept of computation. According to Gross, computational methods do not necessarily represent an optimized version of informal reasoning, rather they are best understood as cognitive tools that can support, extend, and also transform human cognition. In this view, an analysis of computational methods as tools of formal reasoning allows for an analysis of the differences between human and machine-aided cognition and of how they interact in scientific practice.

Emanuele Ratti, in his *Phronesis and Automated Science: The Case of Machine Learning and Biology*, supports the thesis that, since Machine Learning is not independent from human beings, as it is often claimed, it cannot form the basis of automated science. Indeed, although usually computer scientists conceive of their work as being a case of Aristotle's *poiesis* perfected by *techne*, which can be reduced to a set of rules, Ratti argues that there are cases where at each level of computational analysis, more than just *poiesis* and *techne* is required for Machine Learning to work. In this view, biologists need to cultivate something analogous to *phronesis*, which cannot be automated.

A Protocol for Model Validation and Causal Inference from Computer Simulation, by Barbara Osimani and Roland Poellinger, aims to fill in a gap, namely to give a clear formal analysis of computational modelling in systems biology, which is still lacking. To this end, they present a theoretical scheme, which is able to visualize the development of a computer simulation, explicate the relation between different key concepts in the simulation, and trace the epistemological dynamics of model validation. To illustrate such conceptual scheme, they use as a case study the discovery of the functional properties of a protein, E-cadherin, which seems to have a key role in metastatic processes.

Eric Winsberg's *Can Models Have Skill?* aims to determine whether the idea that a model has skill is a step in the direction toward post-human science by focusing on

climate science. Winsberg considers the paradigm of verification and validation which comes from engineering and shows that this paradigm is unsuitable for climate science. He argues that when one deals with models of complex non-linear systems, the best one can find to justify such models is the modeler's explanation to his peers of why it was rational to use a certain approximation technique to solve a particular problem for some specific and contextual reasons. And this shows that science will probably remain human.

In *Virtually Extending the Bodies with (Health) Technologies*, Francesco Bianchini suggests an analogy between the extended mind thesis and a so-called extended body thesis, with particular respect to new technologies connected to health care. According to Bianchini, if one accepts the three main principles which characterize what extends the mind to make it something cognitive, one might wonder whether similar principles are valid for a new vision of the body, which is extended by interactive health technologies. In this perspective, boundaries between what cognition is and what is not could change, as well as boundaries between mind and body could become more blurred.

Finally, the third part of the book, *Automated Science and Human Values*, addresses some relevant ethical issues related to the automation of science and to the scientific endeavor more generally. This part consists of four chapters.

The first chapter of this part is Benjamin Hurlbut's *Behold the Man: Figuring the Human in the Development of Biotechnology*. Accounts of what the human is in debates about biotechnology have mostly focused on the human as an object of technological intervention and control. But, according to Hurlbut, the human being is not separated from the ways of being human together. So, Hurlbut explores the question of whether science will remain human by taking the political conditions of possibility for asking that question as an object of analysis, attending to the ways those political conditions have been transformed in conjunction with the development of biological sciences.

The chapter *The Dehumanization of Technoscience* by Alfredo Marcos addresses the problem of the so-called dehumanization of technoscience, both at the time of its production and at the time of its application. The causes of this twofold dehumanization are found in an oversimplified ontology and in an erratic anthropology, swinging between nihilism and radical naturalism. As an alternative to such perspective, Marcos proposes a pluralistic ontology and an anthropology of Aristotelian inspiration. In this perspective, technoscience becomes valuable and meaningful when it is part of a wider human horizon.

Christopher Tollefsen, in *What is 'Good Science'?*, approaches the question concerning the human character of science from a rather different perspective. In his

view, it is not only automation which threatens the human character of science. Recent controversies over scientific endeavors led some commentators to assert the impropriety of imposing moral limits on scientific inquiry. According to Tollefsen, those claims aim to minimize the human character of science too, since there is instead a deep relationship between good science and morality, which he analyses along three axes, namely the external ethics of science, the social ethics of science, and the internal ethics of science.

In the final chapter, *Cultivating Humanity in Bio- and Artificial Sciences*, Mariachiara Tallacchini offers some reflections on Hurlbut's, Marcos', and Tollefsen's different approaches to the human characters of science presented in previous chapters of this Part. She underlines how, although the accounts of the potential threats to the humanness of science proposed in those chapters follow different trajectories, similar or complementary arguments run across their narratives, revealing that critical voices toward a potential loss of humanity through technology come from multiple and pluralistic perspectives, which are often under-represented in the public debate.

Overall, the contributed chapters in one way or another underline that *something* is missing in the view that science can be made a completely human-independent endeavor, and that philosophical reflection is required nowadays in order to reinforce our understanding of science itself. As it often happens in history, apparent threats turn into constructive challenges both at the individual and collective level. The initial question "Will science remain human?" should thus be reframed by deepening the aspects that humanly characterize scientific praxis, knowledge and understanding. We think that opening up and framing the issue in a wider context of debate is one of the valuable contributions of *this volume* that offers important clue for further researches.

References

Allen, J.F. (2001). In Silico Veritas. Data-Mining and Automated Discovery: The Truth Is in There. *EMBO reports*, 2: 542–544.

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine*, 23 June.

Colton, S. (2002). *Automated Theory Formation in Pure Mathematics*. London: Springer-Verlag.

Gillies, D. (1996). *Artificial Intelligence and Scientific Method*. Oxford, Oxford University Press.

Humphreys, P. (2004). *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*. Oxford: Oxford University Press.

Ioannidis, J.P.A. (2016). Why Most Clinical Research Is Not Useful. *PLOS Medicine*, 13(6): e1002049.

King, R.D. et al. (2009). The Automation of Science (2009). *Science*, 324: 85–89.

Nickles, T. (2018). Alien Reasoning: Is a Major Change in Scientific Research Underway? *Topoi. An International Review of Philosophy*, DOI: 10.1007/s11245-018-9557-1.

Soto, A.M., Longo, G., Noble, D. (eds.) (2016). From the Century of the Genome to the Century of the Organism: New Theoretical Approaches. *Progress in Biophysics and Molecular Biology*, Special Issue, 122(1): 1–82.

Sparkes, A. et al. (2010). Towards Robot Scientists for Autonomous Scientific Discovery. *Automated Experimentation*, 2:1, DOI: 10.1186/1759-4499-2-1.

Sterpetti, F., Bertolaso, M. (2018). The Pursuit of Knowledge and the Problem of the Unconceived Alternatives. *Topoi. An International Review of Philosophy*, DOI: 10.1007/s11245-018-9551-7.

Sweeney, P. (2017). Automated Science as a Vision for AI. *Medium*, available at: https://medium.com/inventing-intelligent-machines/ai-bloatware-why-the-popular-vision-of-ai-is-misleading-ca98c0680f4.