

Introduction

In Silico Modeling: the Human Factor

Marta Bertolaso[†]
m.bertolaso@unicampus.it

Miles MacLeod[‡]
m.a.j.macleod@utwente.nl

Technology is playing an increasingly larger role in biological research and investigation. We are building ever more sophisticated experimental systems which can sample data at rates well beyond anything possible using traditional methods. To analyse this data computational systems do the bulk of the work. Mathematical and computational modelling methods are being applied to large scale systems. These methods can perform analysis and derive information from such systems even with only incomplete information and partial pictures of networks, which is beyond the powers of what experimental work, or indeed ordinary unaided human cognition, could itself achieve.

While human and economic investments in technology development and adoption are growing fast (supported by pressure from many stakeholders) some observers have begun to ask what will all this mean for biology as a science. Undoubtedly, the future of biology is as a technoscience, in which technical and engineering expertise are as important as biological knowledge and experimental skill. As such many of the practices and cultures that have characterized 20th century biology may be supplanted by more automated and algorithmic machine-driven processes. But what can we really expect from technology? How effective will it be and what impact will it have on biological knowledge? How will the role of scientists as human beings be transformed by this epochal transformation? How autonomous will the role of technology be with respect to human contributions in driving research? In sum, how does this human-technology partnership work? Are there any risks or negative drifts that we can foresee and try to counter?

[†] FAST Institute of Philosophy of Scientific and Technological Practice, University Campus Bio-Medico of Rome, Italy.

[‡] Department of Philosophy, University of Twente, The Netherlands.

In this Special Issue we try to lay some foundations for answering these questions by focusing on *in silico* models. *In silico* stands for ‘computational’. Historically, the term *in silico* has played the rhetorical function of giving computational models and simulations the same scientific dignity as *in vitro* and *in vivo* experiments. The same reinforcing function is exploited today, although there is no doubt that *in silico* models are rapidly advancing with new experimental and analytical tools that generate information-rich, high-throughput biological data. For example, they can be used to study complex diseases such as cancer which involve multiple types of biological interactions across diverse physical, temporal, and biological scales. Edelman et al. (2010) express a widespread opinion when they affirm that “Even though cancer has been among the most-studied human diseases using systems approaches, significant challenges remain before the enormous potential of *in silico* cancer biology can be fully realized”. Statistical inferences and models provide an important means for data fitting and discovery; mechanistic knowledge can be reached and used in unprecedented ways; agent-based models enable modeling of cancer cell populations and tumor progression. The promise of *in silico* models is to produce information that can then be used in diagnosis, prognosis and therapy in the clinical setting.

We do not yet necessarily have a good understanding of the affordances and limits of our technological approaches. One can evaluate technological trends by studying how well they live up to their own high expectations and promises. At this point there are still worthy doubts and scepticism to be had. Some promised revolutions have not panned out as hoped (see for instance Genome Wide Association Studies), or at least technology proponents are taking longer than anticipated to find the most appropriate ways to use technology to disentangle biological complexity and variability.

Our use of technology for investigation needs constant shaping. Philosophers tend to turn to the conceptions of biological systems that underlie technological practices and ask deeper methodological and ontological questions about how well technologically supported practices are really adapted to pick up reliable and useful biological signals. They attempt to understand how better notions can frame and inform proper technological use. In this vein we can study the philosophical rationale for systems-level approaches to understanding biological systems, the suitability of biological materials to synthetic manipulation, and the conceptual challenges and practical advantages of reformulating biological knowledge in computationally

accessible frameworks. Such questions have been the principal *foci* of both philosophical and scientific discussions.

Another widespread conviction is that our notions of biological complexity, biological causation and biological organization can themselves be refined by the lessons we draw from technological approaches. In practice, *in silico* big data analyses and automatic pattern recognition are becoming routine in the life sciences: they help us to identify what is regular and what is causally relevant. But the idea that ‘knowledge bases’ could found further knowledge and even decisions is being criticized from several points of view, including technical ones. According to Giuliani (2011), the idea that knowledge may build itself by pure accumulation of smallest ordered pieces of knowledge is quite ancient. For Giuliani, the pursuit of the same idea, now endowed with huge computational power, inevitably heads towards error catastrophes and to “the construction of a no longer amendable false science”. There are theorems in statistics for which demonstrate that very large databases must contain arbitrary correlations, and that too much information tends to behave like very little information. These arguments call at the very least for a cautious attitude towards the autonomy and smartness of technology alone.

Our view is that the emphasis on technology should not interrupt the constitutive relationships that in science link the human observers with the natural world they know and understand. Philosophy and science do offer interesting concepts that foreground these constitutive relationships in science. One is ‘sloppiness’ (Transtrum et al. 2015). Another concept is ‘mesosystem’ (Bizzarri et al. 2011, see also Noble 2006).

Sloppy models have their behaviour controlled by a relatively small number of parameter combinations. They lack a great deal of details. Sloppiness, in its apparent roughness, unifies statistical wisdom and physico-chemical wisdom. Statistics teaches that a model with too many details and parameters that need to be estimated rapidly loses predictivity. This is due to the multiplication of errors and, consequently, to an oversensitivity to contingencies. Physics and chemistry discover systems that are composed of levels of organization, each averaging the lower levels and thus limiting the degrees of freedom that are relevant. By measuring collective properties such as pressure, volume and temperature we can make very precise predictions of system behavior, predictions that are unattainable through any perfect knowledge of the trajectories of single molecules. So the good news is that there are levels at

which the ‘simplification work’ is done by the system through self-organization.

The mesoscopic level is where “organizational principles act on the elementary biological units that will become altered, or constrained, by both their mutual interaction and the interaction with the surrounding environment. In this way and in this place is where general organization behaviour emerges and where we expect to meet the elusive concept of complexity” (Bizzarri et. al. 2011, p. 176). Connected to the mesoscopic way of reasoning is the idea of “mesosystem” (between micro and macro), the system where determinism is maximized for the considered problem under study.

Scientific approaches that work are those that identify the correct level of detail, which is almost never the smallest. Finding the right level at which to seek parts and interactions is a typically human art. This epistemological evidence encourages a cautionary take on technology. It seems that science as a human activity requires that we *leave life in charge* (living systems simplify and drive the research); even better, scientific practice requires that we *leave the human-nature relationship in charge* of shaping and driving the research (Bertolaso 2015). There are many things in contemporary biology that cannot be done by hand. But the human hand needs to be in touch with the nature to be known and, in doing this, to govern and lead research and understanding. It is not only a matter of finding better conceptual foundations for our technologies to make them more reliable and autonomous. The stakes are significance (meaning) and truth of knowledge.

After all, technological approaches, even when highly automated are only put into operation in a human context, defined by sets of human scientific practices, human cognition, and social expectations and demands. Each of these things can serve as constraints on the use of technology. Technology does not replace the need for human practices or human interactions. Such practices and interactions have to be transformed or managed, while technologies have to be adapted to meet them, setting up a complex dynamic between the two.

Our image of technologically assisted research is often shaped by certain specific affordances technological approaches are supposed to offer over ordinary practices, such as automation, objectivity, efficiency, power, precision and so on. Characterizing the role of technology by virtue of these alone however risks basing our analysis on overly idealized pictures of what fields like systems biology, synthetic biology or bioinformatics can realistically

contribute. The problem is that we neglect the human dimension of technoscientific research. A lab-on-a-chip device might offer powerful high-throughput measuring systems, but those systems have to be designed to provide information that usefully substitutes current experimental practices in ways that are as reliable to current biologists. Likewise a bio-ontology needs to be interpretable and accessible to minimally computationally-trained experimental biologists in order to have a prominent role in contemporary biology.

As such while all kinds of efficiencies and possibilities come with computational power, and technological precision, the technology needs to be compatible with its human interface; its users and handlers. Exploring these interfaces opens up a wide potential terrain of investigation that covers issues from sociology of science and medicine, cognitive studies of science, ethics, as well as philosophy. This a truly interdisciplinary terrain.

In this special issue we explore the relationship in particular between new *in silico* or computational modelling methods and the human factor or human element upon which the use and viability of such modelling depends. For these purposes we treat computational research and clinical practices as collective activities performed by human cognitive and social agents. We tackle a set of specific issues: 1) the challenge of adapting computational methods to fit human cognitive capacities in productive way that can build insight in complex systems; 2) the corresponding need for educational training that builds the right sets of skills for handling computation; 3) the challenges of integrating computational modelling with existing often well-established experimental practices in ways that can handle both complexity and also disciplinary boundaries; 4) the challenge of creating a framework for the use of computational methods in personalizing or individualizing medicine, which respects both the likely capacities and limitations of these methods, but also assesses and communicates risk responsibly, and takes account of the public's ability to interpret new computerized individual approaches appropriately without distorting its behaviour in counterproductive directions. Our authors and interviewees approach these challenges as philosophers and scientists keen to understand their methodological and practical implications.

In his paper "Heuristic Strategies in Systems Biology" Fridolin Gross analyzes the lasting necessity of heuristics in the computational models of systems biology. Heuristics are the cognitive strategies that restrict or direct search through any very large problem space, to find the best or most optimal

solutions possible. Sometimes systems biology is described as capable of obtaining a more objective point of view than traditional biology thanks to its consideration of all systems parts, its use of statistical tools and realistic models, and its ability to bring computational algorithmic processes to bear in search strategies to consider a much variety of possibilities. Gross argues against this simplistic view by identifying three main kinds of models in systems biology: small models, large models and network approaches. In each case he shows that the reliance on various heuristics to simplify problems into a computationally tractable state demands an inevitable dependency on human judgment. In small models, “even if some of the cognitive procedures are replaced or extended by algorithmic procedures [...] systems biologists usually do not merely aim at computational tractability, but also demand that the simulations of a model can be followed and eventually comprehended, even if the results are counterintuitive.” Large models have limitations that must be made explicit, rather than hidden in the use of automatic procedures. In fact, large models are themselves often used as heuristic tools for discovery. Network models must be interpreted to acquire their biological meaning, and their interpretation depends once again on strong assumptions. As such while systems biology might be less tied to classical heuristics of “decomposition and localization” it still relies on strong assumptions such as simplicity, sequentiality and modularity, that afford the human mind an ability to apply computation to these systems. Gross encourages scientists who use computational methods to thus recognize the Human Factor in their own research, by making their assumptions explicit (as it is required by their own models), by admitting multiple alternative modeling strategies, and by adopting good means of error detection and correction which result from their own heuristic approaches.

The next two contributions tackle the human challenges of formalizing and standardizing biological knowledge. Boniolo and Lanfrancone articulate the many virtues of formal languages and argue that one major transformation of scientific practice made possible by formalization will be the “automatization of deduction”: molecular pathways that are formalized in a standardized way can be processed by computer programs and turned into sequences of theorems that are demonstrated autonomously by a computer, yielding in some molecular predictions. Complementary software may be designed to mine from the biological literature and code the known pathways into the proper formal language, replacing the time-expensive manual mining and coding. In this way,

the authors argue, formal language can form a bridge between simplicity and complexity. In a complex network of parallel and interacting pathways, single processes may be isolated and formalized, then brought back together with more information, highlighting their reciprocal overlaps and connections. This epistemological strategy, which is considered by Boniolo and Lanfrancone non-reductionist, consists in “deconstruct[ing] a very complex building into its compounding bricks, then reasort[ing] these bricks into small modules according to their logical relationships, and, finally, reconstruct[ing] the original complex building by logically connecting those modules”. To make their points, the authors rely on a concrete tool, *Zsyntax*, a language created to improve the field of network biology by representing the molecular pathways that belong to a network as formal deductions. In particular, they show how two interacting pathways in the melanoma network can be approached with *Zsyntax*, and foreshadow both the generality of their example and the possibility of scaling up the operation to whole interactomes.

Federico Boem agrees with Boniolo and Lanfrancone that computational tools “open the possibility to re-compose complexity”. Taking a philosophy of scientific practice perspective, Boem analyzes the new “ways of doing” science emerging around the Gene Ontology (GO) database. GO is a curated interactive database that represents the features of gene products across different species and databases using a controlled vocabulary. GO is an *applied ontology*, which is a concept of computer science. An applied ontology gathers knowledge from different sources and translates it into a common language, creating new semantic levels. In this way, it achieves a unification of information, which is, according to Boem, “different from theoretical unification”. It also serves to standardize some practices. The Human Factor is quite evident in the GO related practices. First of all, terminological choices (for example, a mechanistic description of molecular events) are made in order to “satisfy the desiderata of the scientific community”. In fact, “such a choice - Boem writes - given the scope and the hope for generality of GO, cannot be grounded just on logical consistency and empirical adequacy”. The GO is “a tool of knowledge-capture and representation”. Secondly, despite the standardization of practices, curators must “possess a robust expertise in the related field”. Thirdly, the GO enters scientific practice as an “orienting tool ... through which scientists can map their data on a wider context and then, thanks to this, elaborate new experimental strategies”. For example, through ‘enrichment analysis’ and through the prediction of putative gene functions the

researchers can zoom out from their most specific research interests and decide where to connect, re-use and explore further. As such, “[i]t is not the practice of experimentation that is changing”; what is changing is “the epistemic role of experiments within research”. The most general implication is, for Boem, that “the rise of ontologies within bioinformatics and their impact on the design of research, should not be understood as a shift from experimental practice to the advent of a sort of ‘*in silico* age’ of the life sciences”, but as a modification of the “hierarchy of methods and evidences of research”. The way in which ontology approaches are applied have less to do with those conceptions we have of computation as automated objective processing and more to do with how human agents will conceptualize a role for them within the scope of their own existing practices.

Annamaria Carusi relies on her fieldwork in the ethnography of science to understand how different expertise and specializations are being coordinated in biomedical labs, where “[i]t is not only a matter of models, simulations and experiments becoming hooked up into a system, but of modellers, simulators, experimenters”. The Human Factor is foregrounded by Carusi, who points out that “the environment of *in silico* modelling is populated by people from many disciplines entering into relationships whereby they agree to undertake joint activities, the outcomes of which need to persuade all involved that it is worthwhile continuing on the iterative cycle of research”; in fact, “there is labour involved in establishing how models correspond to experiments, what in the experiments correspond to, a labour that includes within it agreement and disagreement with others”. Like Boem, Carusi thinks that new epistemological categories are needed to capture the innovative scientific practices that are emerging in the ‘*in silico* age’. In particular, Carusi coins the concept of model-simulation-experiment system, abbreviated as MSE-system. Any MSE-system is a combination of experiments, models and simulations with their reciprocal connections and correspondences, that are progressively tuned to each other over time (for example, by developing uniform visual representations) to take the shape of a knowledge producing system. The tuning of the system is mediated by technology and language, and sustained by persuasions, motivations, values. At any time, “[i]t is the system as a whole that investigates the phenomenon or domain”. At large, there is now a socio-cultural-political movement that presses MSE-systems “to become robust enough to sustain medical and clinical decisions and consequences”. In this respect, Carusi analyzes the rhetorical strategy of the Virtual Physiological

Human (VPH) project, which is also the subject of the next two papers in the issue.

Ilaria Malagrino in her paper “*In silico* Clinical Trials: a new dawn in biomedical research?” reconstructs the history of the VPH project since 2005. The project is not yet entirely realized, but has produced several actions, including the EU funded Avicenna project, started in 2013 and concluded last year. Malagrino focuses on the Avicenna project, “a strategy for *in silico* clinical trials”. *In silico* clinical trials (ISCT) are defined as the use of patient-specific models to generate simulated populations on which new biomedical products can be safely tested. Avicenna was aimed at creating a roadmap for *in silico* clinical trials in Europe, at facilitating research-industry partnerships, and at identifying foundational technologies and paradigmatic case studies. Malagrino analyzes in detail the Avicenna Roadmap, published in September 2015. This very dense document puts forward various arguments (economic, ethical, and also epistemological) for why the EU should support *in silico* clinical trials. The document also identifies the main obstacles to the establishment of ISCT, and by doing this it points out the serious limits of the current system of medical devices and drugs development, including the neglect of rare diseases and the barrier to innovation constituted by patent systems. By introducing a vocabulary of new concepts, such as “virtual patient” and “patient specific modelling”, the Roadmap takes a step towards the creation of a new “*in silico*” culture, which serves as a novel object of philosophical and sociological analysis. Malagrino devotes attention to the collaborative writing process that led to the Roadmap, a process that functioned as negotiation and consensus building platform, thus somewhat fulfilling the explicit goals of Avicenna itself: creating a community of thinkers and laying the foundations for a pre-competitive alliance on ISCT in the EU. Malagrino however raises some deeper philosophical questions that still find no satisfactory answers in such culture and language building efforts as the Avicenna Roadmap. “Even if the strategy of *in silico* Clinical Trials - Malagrino writes - comes as a possible bridge to *ad personam* medical treatments, nevertheless it seems to contain many obscure theoretical issues to overcome in order to create the right horizon and framework upon which we can operationalize its promises”. How does an overarching theoretical framework define ‘disease’ in a living system, for it to be decomposable in underlying biological processes that are, in turn, defined in terms of their constituents or targets? Are mechanistic and quantitative models adequate to reproduce the

living response to a drug or device? Is the virtualization of the human a strategy for transforming clinical trials and, more generally, the whole process of drug and medical device development in the EU? Such issues can presumably only be solved through further discursive processes.

Green and Vogt analyze a view of medicine that is shared by the projects seen in the previous papers. Such an idea is ‘P4 medicine’, standing for predictive, preventive, personalized and participatory. Several scientists and stakeholders are proposing that the classic goal of personalized medicine (i.e., the aim to account for those factors that make health and disease specific for each individual) can now be pursued ‘*in silico*’ via data-integration, and ‘*in socio*’ via patient participation in data collection and disease prevention. Green and Vogt, point out that the crux for evaluating P4 medicine is clinical utility, i.e. the overall balance of benefits vs. harms and costs. Historical data about the low utility of preventive medicine demands strong evidence based on clinical utility for P4 medicine. On the other hand, P4 medicine, by lowering the thresholds and expanding the extent of populations under medical attention, runs the risks of medicalization and overdiagnosis. The authors analyze some specific examples, especially the Hundred Person Wellness Project (HPWP) performed in 2014 by the Institute for Systems Biology in Seattle. At least in the context of HPWP, “the P4 medicine preventive strategy seems to define 100% of a population of previously well as in need of medical attention”. As such P4 medicine advocates have proposed criteria that are too weak for prescribing those in need of medical attention. One of those criteria is ‘actionability’: “successful implementation of P4 medicine not only depends on its ability to accurately predict disease (detect very early signs of disease or risk factors), but also on its ability to translate these predictions into meaningful disease-preventive actions”. If the prediction has no associated effective treatment, it is not actionable. But Green and Vogt notice that concepts such as ‘actionable gene variant’ and ‘actionable possibility’ are vague and require more conceptual work. Further one can also identify issues with the ‘*in socio*’ component of P4 medicine, which are under-considered Human Factors in the implementation of personalized medicine. Some issues concern the patients’ use of more and more detailed medical information. People for instance do not seem to react to risk information in the way that P4 proponents presuppose; for example, the majority of the general public seems to overestimate the benefits of screening. These findings are in tension with claims of P4 medicine about leaving choices about testing to individual

patients. The analysis thus shows that “the social aspects of human life – or the human domain – is of crucial importance for discussions of the prospects of preventive medicine”. Other issues concern incorporating social biomarkers like social networks, religious commitments, general social behavior, into the algorithms of P4 medicine, to ideally reach the total “exposome” of all people. If P4 medicine may be seen as “the culmination of a series of increasingly expansive efforts to improve predictive and preventive strategies to deal with the complexity of human biology and clinical practice”, its expansion certainly presents empirical, epistemological and ethical challenges that, until now, have been addressed with insufficient depth and completeness.

Federica Russo explores the philosophical implications of a technology driven development of environmental epidemiology. The rise of molecular epidemiology has determined a broadening of the notion of ‘exposure’, leading to the idea of ‘exposome’ which is now also recognized by the EU in terms of funded projects. Exposome science is based on the search and establishments of ‘biomarkers of exposure’. As we have seen in other papers, epistemological assumptions and choices play a great role: “While the idea is simple, its implementation is rather complex. There are in fact important and delicate design issues. For instance: when should we make the measurements after exposure? Or, how can we minimise false positive and false negatives?”. Technology (omics, sensors and smartphones, and statistical programs) not only analyse immense data sets: they *produce* them. Understanding the poietic role of technology requires getting back to a classic philosophical debate. Between a ‘subordinate view’ (that maintains a separation between *physis* and *techne* and a priority of either one) and an ‘instrumental view’ (a sort of mediated realism) Russo proposes a third view: technology creates novel “distributed” facts, but the scientist intervenes in constructing them, extrapolating their meaning, and theorising about them.

The Special Issue ends with an interview to three scientists who are directly involved in the implementation and development of *in silico* techniques: Matteo Cerri, neurophysiologist; Markus Reiterer, strategist for Modelling and Simulation (M&S) at Medtronic Inc., one of the major producers of biomedical devices in the world; and Marco Viceconti, among the main leaders of the Virtual Physiological Human initiative and of the Avicenna project and head of the INSIGNEO Institute for *in silico* Medicine, The University of Sheffield. Their CVs and current positions are presented, along with some examples of what do we talk about when we mention *in silico* medicine. The three scientists

discuss how scientific practice is changing from their points of view; what do we have to expect from the constant increase of computational power with respect to our understanding of biological complexity; what are the relationships between research setting and clinical setting; how are risks and responsibilities redistributed in face of the diffusion of *in silico* medicine; and whether and how will the training of scientists need to change to create researchers capable of responding to all these issues.

We provide only a small sampling of the many varied Human Factors that are playing a role in the development of *in silico* modeling methods, many of which are objectively as important as any more computational or technical issue that researchers have to face. By raising them here we hope to show just how much the implementation of technology in biology is a human question. If philosophers and others are to contribute to the technological age in biology, then we should benefit from a much deeper understanding of science and scientific practice that situates technology in its human contexts.

The best time has come to relax specialism and to build a new (but ancient) kind of scientist who knows the foundations and the fundamentals, wise enough to be able to tackle different scientific issues. Complex systems – everybody's object of research nowadays – call for a peculiar style of work, a good training in statistics and multidimensional data analysis, chemistry and physics, and the good sense. The scientist, aware of their role in scientific practice and in the society, must be first of all a well-educated human. This perspective will also help understand and refine the role of technology in biological research, and in the case of medical research, in clinical application.

ACKNOWLEDGMENTS

The editors thank Dr. Emanuele Serrelli for his editorial assistance in realizing this monographic issue.

REFERENCES

Bertolaso, M. (ed.), 2015. *The Future of Scientific Practice: 'Bio-Techno-Logos'*, Pickering & Chatto Publishers, London

- Bizzarri, M., Giuliani, A., Cucina, A., D'Anselmi, F., Soto, A.M., Sonnenschein, C. 2011. Fractal Analysis in a Systems Biology Approach to Cancer. *Seminars in Cancer Biology* 21:175-82.
- Edelman, L.B., Eddy, J. a & Price, N.D., 2009. In silico models of cancer. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 2(4), pp.438-459.
- Giuliani A. (2011) Science as Theater: too obvious to be appreciated. *Topoi* 30 (2): 165-171.
- Noble, D. 2006. *The music of life: biology beyond the genome*. Oxford: Oxford University Press.
- Transtrum, M. K., Machta, B. B., Brown, K. S., Daniels, B. C., Myers, C. R., & Sethna, J. P. (2015). Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *The Journal of Chemical Physics*, 143(1), 010901).

