Draft paper for the symposium *Mechanism Meets Big Data: Different Strategies for Machine Learning in Cancer Research* to be held at *PSA2018: The 26th Biennial Meeting of the Philosophy of Science Association* (Seattle, WA; 1-4 Nov 2018).

MECHANISTIC MODELS AND THE EXPLANATORY LIMITS OF MACHINE LEARNING

Emanuele Ratti[1], University of Notre Dame

Ezequiel López-Rubio, Universidad Nacional de Educación a Distancia, University of Málaga

Abstract

We argue that mechanistic models elaborated by machine learning cannot be explanatory by discussing the relation between mechanistic models, explanation and the notion of intelligibility of models. We show that the ability of biologists to understand the model that they work with (i.e. intelligibility) severely constrains their capacity of turning the model into an explanatory model. The more a mechanistic model is complex (i.e. it includes an increasing number of components), the less explanatory it will be. Since machine learning increases its performances when more components are added, then it generates models which are not intelligible, and hence not explanatory.

1. INTRODUCTION

Due to its data-intensive turn, molecular biology is increasingly making use of machine learning (ML) methodologies. ML is the study of generalizable extraction of patterns from data sets starting from a problem. A problem here is defined as a given set of input variables, a set of outputs which have to be calculated, and a sample (previously input-output pairs already observed). ML calculates a quantitative relation between inputs and outputs in terms of a predictive model by learning from an already structured set of input-output pairs. ML is expected to increase its performances when the complexity of data sets increase, where complexity refers to the number of input variables and the number of samples. Due to this capacity to handle complexity, practitioners think that ML is potentially able to deal with biological systems at the macromolecular level, which are notoriously complex. The development of ML has been proven useful not just for the

[1] mnl.ratti@gmail.com

complexity of biological systems *per se*, but also because biologists now are able to generate an astonishingly amount of data. However, we claim that the ability of ML to deal with complex systems and big data comes at a price; *the more ML can model complex data sets, the less biologists will be able to explain phenomena in a mechanistic sense*.

The structure of the paper is as follows. In Section 2, we discuss mechanistic models in biology, and we emphasize a surprising connection between explanation and model complexity. By adapting de Regt's notion of pragmatic understanding (2017) in the present context, we claim that if a how-possibly mechanistic model can become explanatory, then it must be intelligible to the modeler (Section 2.2, 2.3 and 2.4). Intelligibility is the ability to perform precise and successful material manipulations on the basis of the information provided by the model about its components. The results of these manipulations are fundamental to recompose the causal structure of a mechanism out of a list of causally relevant entities. Like a recipe, the model must provide instructions to 'build' the phenomenon, and causal organization is fundamental in this respect. If a model is opaque to these organizational aspects, then no mechanistic explanations can be elaborated. By drawing on studies in cognitive psychology, we show that the more the number of components in a model increases (the more the model is complex), the less the model is intelligible, and hence the less an explanation can be elaborated.

Next, we briefly introduce ML (Section 3). As an example of ML application to biology, we analyze an algorithm called PARADIGM (Vaske et al 2010), which is used in biomedicine to predict clinical outcomes from molecular data (Section 3.1). This algorithm predicts the activities of genetic pathways from multiple genome-scale measurements on a single patient by integrating information on pathways from different databases. By discussing the technical aspects of this algorithm, we will show how the algorithm generates models which are more accurate as the number of variables included in the model increases. By variables, here we mean biological entities included in the model and the interactions between them, since those entities are modeled by variables in PARADIGM.

In Section 4 we will put together the results of Section 2 and 3. While performing complex localizations more accurately, we argue that an algorithm like PARADIGM makes mechanistic models so complex (in terms of the number of model components) that no explanation can be constructed. In other words, ML applied to molecular biology undermines biologists' explanatory abilities.

2. COMPLEXITY AND EXPLANATIONS IN BIOLOGY

The use of machine learning has important consequences for the explanatory dimension of molecular biology. Algorithms like PARADIGM, while providing increasingly accurate localizations, challenge the explanatory abilities of molecular biologists, especially if we assume the account of explanation of the so-called mechanistic philosophy (Craver and Darden 2013; Craver 2007; Glennan 2017). In order to see how, we need to introduce the notion of mechanistic explanation, and its connection with the notion of intelligibility (de Regt 2017).

2.1 Mechanistic explanations

Molecular biology's aim is to explain how phenomena are produced and/or maintained by the organization instantiated by macromolecules. Such explanations take the form of mechanistic descriptions of these dynamics. As Glennan (2017) succinctly emphasizes, mechanistic models (often in the form of diagrams complemented by linguistic descriptions) are vehicles for mechanistic explanations. Such explanations show how a phenomenon is produced/maintained and constituted by a mechanism – mechanistic models explain by explaining *how*. As Glennan and others have noticed, a mechanistic description of a phenomenon looks like what in historical narrative is called *causal narrative*, in the sense that it "describes sequences of events (which will typically be entities acting and interacting), and shows how their arrangement in space and time brought about some outcome" (Glennan 2017, p 83). The main idea is that we take a set of entities and activities to be causally relevant to a phenomenon, and we explain the phenomenon by showing how a sequence of events involving the interactions of the selected entities produces and/or maintains the explanandum. In epistemic terms, it is a

matter of showing a chain of inferences that holds between the components of a model (e.g. biological entities). Consider for instance the phenomenon of restriction in certain bacteria and archaea (Figure 1). This phenomenon has been explained in terms of certain entities (e.g. restriction and modification enzymes) and activities (e.g. methylation). Anytime a bacteriophage invades one of these bacteria or archaea (from now on *host cells*), host cells stimulate the production of two types of enzymes, i.e. a restriction enzyme and a modification enzyme. The restriction enzyme is designed to recognize and cut specific DNA sequences. Such sequences, for reasons we will not expose here[2], are to be found in the invading phages and/or viruses. Hence, the restriction enzyme destroys the invading entities by cutting their DNA. However, the restriction enzyme is not able to distinguish between the invading DNA and the DNA of the host cell. Here the modification enzyme helps, by methylating the DNA of the host cell at specific sequences (the same that the restriction enzyme cuts), thereby preventing the restriction enzyme to destroy the DNA of the host cell. The explanation of the phenomenon of restriction is in terms of a narrative explaining how certain entities and processes contribute to the production of the phenomenon under investigation. The inferences take place by thinking about the characteristics of the entities involved, and how the whole functioning of the system can be recomposed from entities themselves.

[2] See for instance (Ratti 2018)

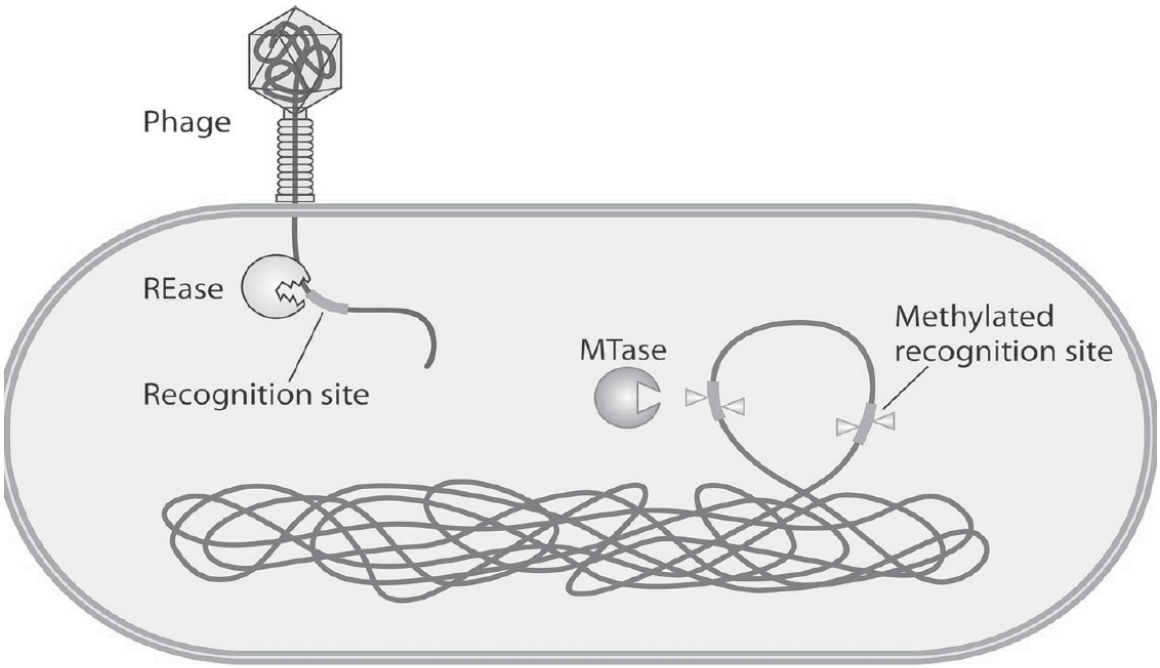


Figure 1. Mechanistic model of restriction. A phage enters a bacterium cell and sequences of its DNA are cleaved by a restriction enzyme (REase). Simultaneously, a modification enzyme (MTase) methylates a specific sequence in the DNA of host so that the restriction enzyme does not cleave the genome of the host too. Original figure taken from (Vasu and Nagaraja 2013).

2.2. Complexity of mechanistic models

Despite the voluminous literature on mechanistic explanation, there is a connection between models, *in fieri* explanations and the modeler that has not been properly characterized. In particular, mechanistic models should be intelligible to modelers in order to be turned into complete explanations. Craver noticed something like that when he states that his ideal of completeness of a mechanistic description (in terms of molecular details) should not be taken literary, but completeness always refer to the particular explanatory context one is considering. The reason why literary completeness is unattainable is because complete models will be of *no use* and completely *obscure* to modelers; "such descriptions would include so many potential factors that they would be *unwieldy for the purpose of prediction and control and utterly unilluminating to human beings*" (2006, p 360, emphasis added).

We rephrase Craver's intuitions by saying that *how-possibly models cannot be turned into adequate explanations if they are too complex*. We define complexity as *a function of the number of entities and activities (i.e. components of the model) that have*

to be coordinated in an organizational structure in the sense specified by mechanistic philosophers. This means that no agent can organize the entities and/or activities localized by highly complex models in a narration that rightly depicts the organizational structure of the *explanandum*. Therefore, very complex models which are very good in localization cannot be easily turned into explanations. Let us show why complex models cannot be turned into explanatory models in the mechanistic context.

2.3 Intelligibility of mechanistic models

The idea that agents cannot turn highly complex mechanistic models into explanations can be made more precise by appealing to the notion of *intelligibility* (de Regt 2017).

By following the framework of models as mediators (Morgan and Morrison 1999), de Regt argues that models are the way theories are applied to reality. Similar to Giere (2010), de Regt thinks that theories provide principles which are then articulated in the form of models to explain phenomena; "[t]he function of a model is to represent the target system in such a way that the theory can be applied to it" (2017, p 34). He assumes a broad meaning of explanation, in the sense that explanations are arguments, namely attempts to "answer the question of why a particular phenomenon occurs or a situation obtains (…) by presenting a systematic line of reasoning that connects it with other accepted items of knowledge" (2017, p 25). *Ça va sans dire*, arguments of the sort are not limited to linguistic items[3]. On this basis, de Regt's main thesis is that a *condition sine qua non* to elaborate an explanation is that the theory from which it is derived must be intelligible.

In de Regt's view, the intelligibility of a theory (*for scientists*) is "[t]he value that scientists attribute to the cluster of virtues (…) that facilitate the use of the theory for the construction of models" (p 593). This is because an important aspect of obtaining explanations is to derive models from theories, and to do that a scientist must use the theories. Therefore, if a theory possesses certain characteristics that make it easier to be used by a scientist, then the same scientist will be in principle more successful in deriving explanatory models. In (2015) de Regt extends this idea also to models in the sense that "understanding consists in being able to use and manipulate the model in order to make

[3] Mechanistic explanations are arguments, though not of a logical type

inferences about the system, to predict and control its behavior" (2015, p 3791). If for some reasons models and theories are not intelligible (to us), then we will not be able to develop an explanation, because we would not know how to use models or theories to elaborate one.

This idea of intelligibility of models and its tight connection with scientific explanation, can be straightforwardly extended to mechanistic models. Intelligibility of mechanistic models is defined by the way we *successfully* use them to explain phenomena. But how do we use models (mechanistic models in particular), and for what? Please keep in mind that whatever we do with mechanistic models, it is with explanatory aims in mind. Anything from predicting, manipulating, abstracting, etc is because we want an explanation. This is a view shared both by mechanistic philosophers but by de Regt as well, whose analysis of intelligibility is in explanatory terms.

First, highly abstract models can be used to build more specific models, as in the case of schema (Machamer et al 2000; Levy 2014). A schema is "a truncated abstract description of a mechanism that can be filled with descriptions of known component parts and activities" (Machamer et al 2000, p 16). For instance, consider the model of transcription. This model can be highly abstract where 'gene' stands for any gene, and 'transcription factor' stands for any transcription factor. However, we can instantiate such a schema in a particular experimental context by specifying which gene and which transcription factors are involved. The idea is that biologists, depending on the specific context they are operating, can instantiate experiments to find out which particular gene or transcription factor is involved in producing a phenomenon at a given time.

Next, mechanistic models can be used in the context of the *build-it test* (Craver and Darden 2013) with confirmatory goals in mind. Since mechanistic explanations may be understood as recipes for construction, and since recipes provide instructions to use a set of ingredients and instruments to produce something (e.g. a cake), then mechanistic models provide instructions to build a phenomenon or instructions to modify it in controlled ways because, after all, they tell us about the internal division of labor between entities causally relevant to producing or maintaining phenomena. This is in essence the build-it test as a confirmation tool; by modifying an experimental system on the basis of the 'instructions' provided by the model that allegedly explains such a phenomenon, we

get hints as to how the model is explanatory. If the hypothesized modifications produce in the 'real-world' the consequences we have predicted on the basis of the model, then the explanatory adequacy of the model is corroborated. The more the modifications suggested are precise, the more explanatory the model will be[4]. A first lesson we can draw is that *if a mechanistic model is explanatory, then it is also intelligible*, because it is included in the features of being explanatory mechanistically the fact that we can use the model to perform a build-it test.

The build-it test is also useful as a *tool to develop* explanations. Consider again the case of restriction in bacteria and a how-possibly model of this phenomenon based on a few observations. Let's say that we have noticed that when phages or viruses are unable to grow in specific bacteria, such bacteria also produce two types of enzymes. We know that the enzymes, the invading phages/viruses and restriction are correlated. The basic model will be as follows; anytime a phage or a virus invade a bacterium, these enzymes are produced, and hence the immune system of the bacterium must be related to these enzymes. We start then to instantiate experiments on the basis of this simple model. Such a model suggests that these enzymes must do something to the invading entities, but that somehow modify the host cell as well. Therefore, the build-it test would consist in a set of experiments to stimulate and/or inhibit these entities to develop our ideas about the nature of their causal relevance and their internal division of labor. *In fieri* mechanistic models suggest a range of instructions to 'build' or 'maintain' phenomena. These instructions are used to instantiate experiments to refine the model and make it explanatory. This is an example of what Bechtel and Richardson would call *complex localization* (2010, Chapter 6), and it is complex because the strategy used to explain the behavior of a system (immune system of host cells) is heavily constrained by empirical results of lower-levels. The how-possibly model affords a series of actions leading to a case of complex localization, when "constraints are imposed, whether empirical or theoretical, they can serve simultaneously to vindicate the initial localization and to develop it into a full-blooded mechanistic explanation" (Bechtel and Richardson 2010, p 125). Therefore, *if a how-possibly model can be turned into an explanatory model, then it*

[4] Please note that such a test, when involving adequate mechanistic explanations, is also the preferred way to teach students in text books, or also a way to provide instructions to reproduce the results of a peer-reviewed article

is intelligible, because the way we turn it into an explanatory model is by instantiating build-it tests.

A mechanistic model is therefore intelligible either when (a) it is a schema and we can instantiate such a model in specific contexts, or (b) when it affords a series of built-it test which are used either to corroborate its explanatory adequacy, or to make it explanatory. About (b), it should be noted that if we consider a mechanistic model as a narrative, then the model will be composed of a series of steps which influence each other in various ways. *Being able to use a model means being able to anticipate what would happen to other steps if I modify one step in particular*. This is not a yes/no thing. The model of restriction-modification systems is highly intelligible, because I know that if I prevent the production of modification enzymes I simultaneously realize that the restriction enzyme will destroy the DNA of the host cell. However, more detailed models will be less intelligible, because it would be difficult to simultaneously anticipate what would happen at each step by modifying a step in particular.

2.4 Recomposing mechanisms and intelligibility

In the mechanistic literature, the process of developing an explanatory model out of a catalogue of entities that are likely to be causally relevant to a phenomenon is called *recomposition of a mechanism* and it usually happens after a series of localization steps.

To recompose a mechanism, a modeler must be able to identify causally relevant entities and their internal division of labor. The idea is not just to 'divide up' a given phenomenon in tasks, but also a given task in subtasks interacting in the overall phenomenon, as it happens in complex localization (Bechtel and Richardson 2010). In the simplest case, researchers assume linear interactions between tasks, but there may be also non-linear or more complex type of interactions.

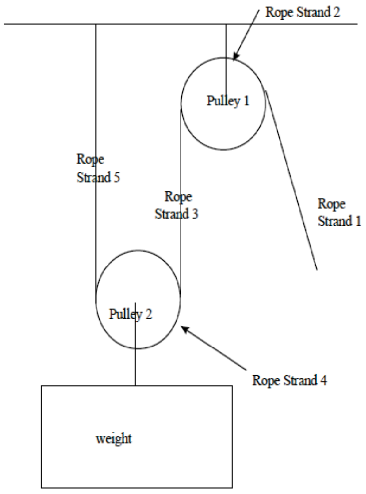
These reasoning strategies are usually implemented by thinking about these dynamics with the aid of *diagrams*. Diagrammatic representations usually involve boxes standing for entities (such as genes, proteins, etc) and arrows standing for processes of various sorts (phosphorylation, methylation, binding, releasing, etc). Therefore, biologists recompose mechanisms as mechanistic explanations by thinking about these diagrams,

and they instantiate experiments (i.e. built-it test) exactly on the basis of such diagrammatic reasoning.

Cognitive psychology and studies of scientific cognition have extensively investigated the processes of diagrammatic reasoning (Hegarty 2000; 2004; Nersessian 2008). Moreover, empirical studies have emphasized the role of diagrams in learning and reasoning in molecular biology (Kindfield 1998; Trujillo 2015). In these studies, diagrammatic reasoning is understood as a "task that involves inferring the behavior of a mechanical system from a visual-spatial representation" (Hegarty 2000, p 194). Hegarty refers to this process as *mental animation*, while Nersessian (2008) thinks about this as an instantiation of *mental modelling*. This is analogous to thinking about mechanistic models as narratives, namely being able to infer how a course of events, decomposed into steps, may change if we change one step in particular. Mental animation is a process of complex visual-spatial inference. Limits and capabilities of humans in such tasks depend on the cognitive architecture of human mind[5]. What Hegarty has found is that mental animation is *piecemeal*, in the sense that human mind does not animate the components of a diagram in parallel, but rather infer the motion of components *one by one*. This strategy has a straightforward consequence; in order to proceed with animating components, we should store intermediate results of inferences drawn on previous components. Due to the limitations of working memory (WM), people usually store such information on external displays. Hegarty has provided evidence that diagrammatic reasoning is bounded to WM abilities. The more we proceed in inferring animation on later components, the more the inferences on earlier components degrade (see for instance Figure 2); "as more components of the system are 'read into' spatial working memory, the activation of all items is degraded, so that when later components are in, there is not enough activation of the later components to infer their motion" (Hegarty 2000, p 201).

[5] On this, I rely on the framework assumed by the cognitive-load theory (Paas et al 2010)

a



b

Cycle	Action Taken
1	Rope Strand 1 is inspected (simulation is given knowledge that this rope strand is being pulled)
2	Rope Strand 1 is animated (i.e. its direction of motion is inferred)
3	Rope Strand 2 is inspected
4	Rope Strand 2 is animated
5	Rope Strand 3 is inspected Pulley 1 is inspecte
6	Rope Strand 3 is animated Pulley 1 is animated
7	Rope Strand 4 is inspected
8	Rope Strand 4 is animated
9	Rope Strand 5 is inspected Pulley 2 is inspected
10	Rotation of Pulley 2 is inferred Translation of Pulley 2 is inferred
11	Weight is inspected
12	Weight is animated

Figure 2. (a) Example of diagram of a simple pulley system that can be mentally animated (b) Description of typical actions that can be one by one to animate the pulley system. Both figures taken from Hergaty (2000)

The actual limit of our cognitive architecture on this respect may be debated, and it is an empirical issue. The important point is that *no matter our external displays*, for very large systems (such as Figure 3) it is very unlikely that human cognition will be able to process all information about elements interactivity. This is because by animating components one-by-one, even if we use sophisticated instruments such computer simulations, still inferences on earlier components will degrade. This means that build-it tests will be very ineffective, if not impossible. In terms of narratives, recipes and mechanistic models, this means that for large mechanistic diagrams with many model components, no human would be able to anticipate the consequences of modifying a step in the model for all the other steps of the model, even if a computer simulation shows that the phenomenon can be possibly produced by the complex model. The computer simulation may highlight certain aspects (as Bechtel in 2016 notes), but the model is not intelligible in the sense required by mechanistic philosophy. *If the model is not intelligible in this way, then it cannot be possibly turned into an explanation*.

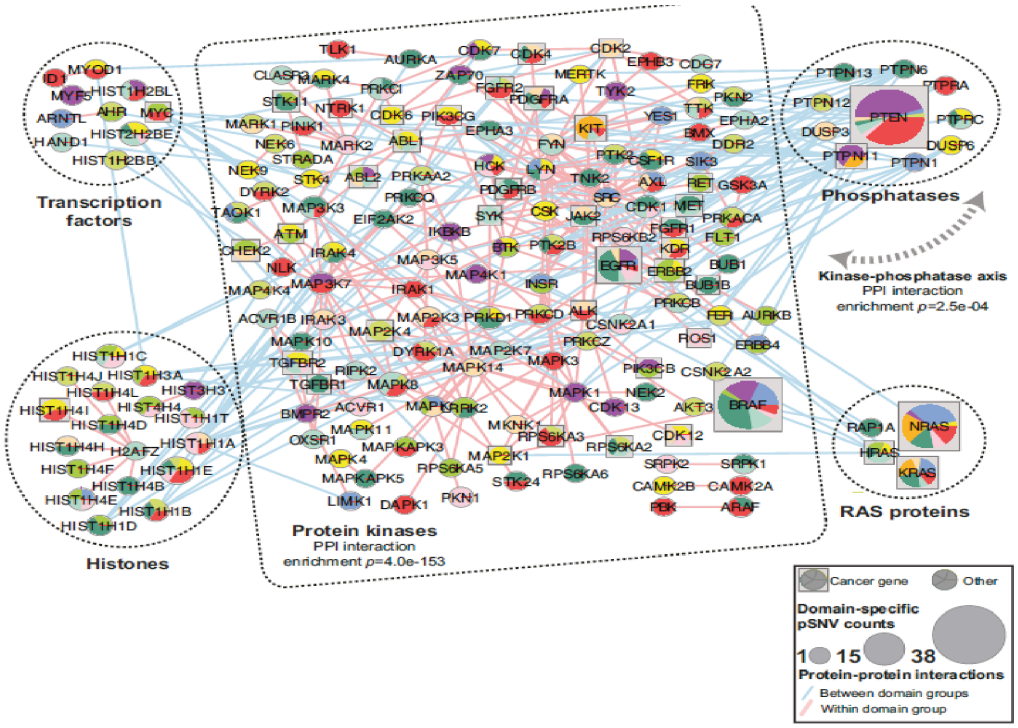


Figure 3. Network of interactions of proteins with significant enrichment of phosphorylation-related single nucleotide variations. Phosphorylation is a central post-translational modification in cancer biology. Authors are not trying to recompose the mechanism that from phosphorylated proteins (nodes) lead to a tumor phenotype, but rather to identify the magnitude of the impact of this process on cancer genes. Figure taken from (Reimand et al 2013)

The results of Hegarty's research suggest that when mechanistic models are concerned, strategies of localization are effective (in terms of explanatory potential) only when a limited number of model components are actually identified. The number may increase if we use computer simulations. However, for very large amounts of model components (such as Figure 3) recomposition is just impossible for humans, because inferences on the role of components in the causal division of labor of a phenomenon will degrade to make place for inferences about other components. This of course holds only if we have explanatory aims in mind.

To summarize, in section 2 we have made three claims:

1. If a how-possibly model can be turned into an explanation, then it is intelligible
2. If a model is not intelligible, then it cannot become explanatory
3. Complex models are a class of non-intelligible models

3. MACHINE LEARNING AND LOCALIZATIONS

Machine learning (ML) is a subfield of computer science which studies the design of computing machinery that improves its performance as it learns from its environment. A ML algorithm extracts knowledge from the input data, so that it can give better solutions to the problem that it is meant to solve. This learning process usually involves the automatic construction and refinement of a model of the incoming data. In ML terminology, a model is an information structure which is stored in the computer memory and manipulated by the algorithm.

As mentioned before, the concept of 'problem' in ML has a specific meaning which is different from other fields of science. A ML problem is defined by a set of input variables, a set of output variables, and a collection of samples which are input-output pairs. Solving a problem here means finding a quantitative relation between inputs and outputs in the form of a predictive model, in the sense that the algorithm will be used to produce a certain output given the presence of a specific input.

3.1 The PARADIGM algorithm

ML has been applied in the molecular sciences in many ways (Libbrecht and Noble 2015). Especially in cancer research[6], computer scientists have created and trained a great deal of algorithms in order to identify entities that are likely to be involved in the development of tumors, how they interact, to predict phenotypes, to recognize crucial sequences, etc (see for instance Leung et al 2016).

As a topical example of ML applied to biology, we introduce an algorithm called PARADIGM (Vaske et al 2010). This algorithm is used to infer how genetic changes in a patient influence or disrupt important genetic pathways underlying cancer progression. This is important because there is empirical evidence that "when patients harbor genomic alterations or aberrant expression in different genes, these genes often participate in a common pathway" (Vaske et al 2010, p i237). Because pathways are so large and biologists cannot hold in their mind the entities participating in them, PARADIGM integrate several genomic datasets – including datasets about interactions between genes and phenotypic consequences – to infer molecular pathways altered in patients; it predicts

[6] See for instance The Cancer Genome Atlas at https://cancergenome.nih.gov

whether a patient will have specific pathways disrupted given his/her genetic mutations.

The algorithm is based on a simplified model of the cell. Each biological pathway is modeled by a graph. Each graph contains a set of nodes, such that each node represents a cell entity, like a mRNA, a gene or a complex. A node can be only in three states (i.e. activated, normal or deactivated). The connections among nodes are called factors, and they represent the influence of some entities on other entities. It must be noticed that the model does not represent why or how these influences are exerted. Only the sign of the influence, i.e. positive or negative, is specified.

The model specifies how the expected state of an entity must be estimated. The entities which are connected by positive or negative factors to the entity at hand cast votes which are computed by multiplying +1 or –1 by the states of those entities, respectively. In addition to this, there are 'maximum' and 'minimum' connections to cast votes which are the maximum or the minimum of the states of the connected entities, respectively. Overall, the expected state of an entity is computed as the result of combining several votes obtained from the entities which are connected to it. Such a voting procedure can be associated to localizations (i.e. whether a node is activated or not), but hardly to biological explanations.

The states of the entities can be hidden, i.e. they can not be directly measured on the patients, or observable. The states of the hidden variables must be estimated by a probabilistic inference algorithm, which takes into account the states of the observed variables and the factors to estimate the most likely values of the hidden variables. Here it must be pointed out that this algorithm does not yield any explanation about the computed estimation. Moreover, it could be the case that the estimated values are not the most likely ones, since the algorithm does not guarantee that it finds the globally optimum solution.

The size of the model is determined by the number of entities and factors that the scientist wishes to insert. A larger model provides a perspective of the cell processes which contains more elements, and it might yield better predictions. This means that the more components the model has, the better the algorithm will perform. In biological terms, the larger the model, the more precise *complex localizations* the algorithm will identify, in particular by pointing more precisely towards pathways that are likely to be

disrupted in the patient with more information about the state of gene activities, complexes and cellular processes. Importantly, PARADIGM does not infer new genetic interactions, but it just helps identifying those known interaction in a new data set. It is completely supervised, in the sense that "[w]hile it infers hidden quantities (…), it makes no attempt to infer new interactions not already present in an NCI [National Cancer institute database] pathway" (Vaske et al 2010, p i244).

4 COMPLEX MODELS AND MECHANISTIC EXPLANATIONS

Before unwinding our conclusions, let me recall the results of Section 2 very briefly:

1. If a how-possibly model can be turned into an explanation, then it is intelligible[7].
2. If a model is not intelligible, then it cannot become explanatory
3. Complex model (in the sense explained in 2.2) are not intelligible

What does this have to do with PARADIGM? It is important to emphasize what we have pointed out in Section 3.1, namely that an algorithm like PARADIGM is more efficient when working with more components. If we think about models generated by algorithms such as PARADIGM in mechanistic terms, this means that the algorithm provides more precise complex localizations, because more entities that are likely to be causally relevant to a phenomenon are identified, and the information about the probability of a pathway being disrupted in a patient will be more precise. However, the models will be more complex, and they will be decreasingly intelligible. This is because the final model will count an elevated number of components, and recomposing these components into a full-fledged mechanistic explanation of how a tumor is behaving will be cognitively very difficult; the inferences about the behavior of components are not run in parallel, but one by one, and once we proceed in inferring the behavior of a component on the basis of the behavior of another component, other inferences will degrade, as Hegarty's studies have shown. In the ideal situation, PARADIGM will generate unintelligible models:

[7] Remember: A mechanistic model *x* is intelligible to a modeler *y* if *y* can use the information about the components of *x* to instantiate so-called 'build-it test'. Such tests are performed on how-possibly models to turn them into explanatory models by obtaining information on how to recompose a phenomenon (i.e. by showing how a list of biological entities are organized to produce a phenomenon).

4. Algorithms such as PARADIGM generate models which are not intelligible because such models are too complex
5. Because of 2, 3 and 4, complex models generated out of algorithms like PARADIGM cannot become explanations

This means that when we use algorithms such as PARADIGM to cope with the complexity of biological systems, we successfully handle big data sets, but such a mastery comes at a price. Using ML in molecular biology means providing more detailed localizations, but we also lose explanatory power, because no modeler will be able to recompose the mechanism out of a long list of entities.

This implies that, in the mechanistic epistemic horizon, the central role assigned to explanations should be reconsidered when contemporary molecular biosciences are concerned. As Bechtel has also emphasized in the context of computational models in mechanistic research (2016), such tools are useful to show whether some entities are likely to be involved in a particular phenomenon or suggest alternative hypotheses about the relation between certain entities. However, providing fully-fledged mechanistic explanations is another thing. It is the same with algorithms of ML; we identify more entities likely to be involved in a mechanism, we may even find out that entities involved in specific process may be connected with entities involved in other processes (via for instance Gene Ontology enrichments), but we cannot recompose a mechanism out of a list of hundreds of entities. In fact, we come to value different epistemic values, and *explanatory power is not one of them*. This somehow implies also a shift in the way scientific articles are organized; if in 'traditional' molecular biology evidence converges towards the characterization of a single mechanism, in data-intensive biology we make a list of entities that can be involved in a phenomenon, but we do not necessarily connect those entities mechanistically (Alberts 2012). Another strategy (Krogan et al 2015) – though motivated more by biologically rather than cognitive reasons – is to abstract from macromolecular entities and consider only aggregates of them in the form of networks; whether establishing network topology is providing a mechanistic explanation remains an open question.

REFERENCES

Alberts, B. (2012). The End of "Small Science"? *Science*, *337*(September), 1230529.

Bechtel, W. (2016). Using computational models to discover and understand mechanisms. *Studies in History and Philosophy of Science Part A*, *56*, 113–121.

Bechtel, W., & Richardson, R. (2010). *Discovering Complexity - Decomposition and Localization as Strategies in Scientific Research*. Cambridge, Massachusetts, and London, England: The MIT Press.

Craver, C. (2007). *Explaining the Brain - Mechanisms and the Mosaic Unity of Neuroscience*. Oxford University Press.

Craver, C. F. (2006). When mechanistic models explain. *Synthese*, *153*(3), 355–376.

Craver, C., & Darden, L. (2013). *In search of Mechanisms*. Chicago: The University of Chicago Press.

De Regt, H. (2017). *Understanding Scientific Understanding*. Oxford: Oxford University Press.

de Regt, H. W. (2015). Scientific understanding: truth or dare? *Synthese*, *192*(12), 3781–3797. http://doi.org/10.1007/s11229-014-0538-7

Giere, R. N. (2010). An agent-based conception of models and scientific representation. *Synthese*, *172*(2), 269–281.

Glennan, S. (2017). *The New Mechanical Philosophy*. Oxford University Press.

Hegarty, M. (2000). Capacity Limits in Mechanical Reasoning. In M. Anderson, P. Cheng, & V. Haarslev (Eds.), *Diagrams 2000* (pp. 194–206). Springer-Verlag.

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences*, *8*(6), 280–285.

Krogan, N. J., Lippman, S., Agard, D. A., Ashworth, A., & Ideker, T. (2015). The Cancer Cell Map Initiative: Defining the Hallmark Networks of Cancer. *Molecular Cell*, *58*(4), 690–698.

Levy, A. (2014). What was Hodgkin and Huxley's achievement? *British Journal for the Philosophy of Science*, *65*(3), 469–492.

Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, *16*(6), 321–332.

Machamer, P., Darden, L., & Craver, C. (2000). Thinking about Mechanisms. *Philosophy of Science*, (67), 1–25.

Morrison, M., & Morgan, M. (1999). Models as mediating instruments. In M. Morrison & M. Morgan (Eds.), *Models as Mediators*. Cambridge University Press.

Nersessian, N. (2008). *Creating Scientific Concepts*. Cambridge, MA: The MIT Press.

Ratti, E. (2018). "Models of" and "models for": On the relation between mechanistic models and experimental strategies in molecular biology. *British Journal for the Philosophy of Science*.

Reimand, J., Wagih, O., & Bader, G. D. (2013). The mutational landscape of phosphorylation signaling in cancer. *Scientific Reports*, *3*.

Vaske, C. J., Benz, S. C., Sanborn, J. Z., Earl, D., Szeto, C., Zhu, J., … Stuart, J. M. (2010). Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics*, *26*(12), 237–245.

Vasu, K., & Nagaraja, V. (2013). Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense. *Microbiology and Molecular Biology Reviews*, *77*(1), 53–72.