# The Devil in the Data: Machine Learning & the Theory-Free Ideal

Mel Andrews

October 24, 2023

**Abstract**

Philosophers of science have argued that the widespread adoption of the methods of machine learning (ML) will entail radical changes to the variety of epistemic outputs science is capable of producing. Call this the *disruption claim*. This, in turn, rests on a *distinctness claim*, which holds ML to exist on novel epistemic footing relative to classical modelling approaches in virtue of its *atheoreticity*. We describe the operation of ML systems in scientific practice and reveal it to be a necessarily theory-laden exercise. This undercuts claims of epistemic distinctness and, therefore, at least one path to claims of disruption.

## 1 Introduction

The field of artificial intelligence—or, perhaps more accurately, fields, as there have been a number of distinct research communities and bodies of scholarship under this denomination which bear little in common beyond the title—has been beset by waves of popular attention and ebbs and flows of funding since its incipience. It is worth noting that there have been (at least) two epochs of AI hype in the past decades which have compounded together to emerge in the present form of obsession over "automated science." In the late 1990s and early 2000s there occurred an early internet data-mining craze, in which it was discovered that statistical techniques, unleashed on sufficiently large datasets, could reveal complex and meaningful patterns (Rajaraman & Norvig, 1998). Around 2010 there commenced what has since come to be known as the deep learning revolution, in which breakthroughs in neural network approaches to image classification (alexnet) and natural language processing (word2vec) renewed public interest in the methods of machine learning. We appear to be on the cusp of a third wave at present with the ascension of generative modelling fueled by web-scale data.

The prospects of machine learning for science have, in particular, opened wide in the last decade, over which timescale ML techniques were adopted in the Large Hadron Collider at CERN for sorting the significance of particle collision events (Duarte et al., 2018) and DeepMind released its AlphaFold and

AlphaFold 2.0 (Jumper et al., 2021), capable of predicting tertiary and quaternary protein structure from amino acid sequence data, effectively solving one of biology's most complex and enduring open problems. The rapidity and ubiquity of machine learning uptake across all sectors of public life, in particular, science, has sparked an onslaught of speculation concerning its nature and the downstream consequences of its widespread use.

Such speculation has issued from cultural commentators, journalists, and media personalities, from the mathematicians and engineers producing the tools of ML and the scientists deploying them and, of course, from philosophers, in both academic and popular venues. Responses focussed on the epistemic status of ML and its projected impact on science have echoed statements to the effect that machine learning differs radically from prevailing modelling, statistical, or scientific methods in ways that are projected to change the landscape of scientific discovery or the nature of the epistemic fruits of scientific enterprise.

These scholars prophesy a data-driven or ML-driven scientific revolution (Hey, Tansley, Tolle, et al., 2009; Mayer-Schönberger & Cukier, 2013; Anderson, 2008; Spinney, 2022). They foretell an end to hypothesis testing (Anderson, 2008; Mayer-Schönberger & Cukier, 2013; Spinney, 2022). They envision ML methods retiring or else displacing the role of theorising in science (Anderson, 2008; Mayer-Schönberger & Cukier, 2013; Spinney, 2022; Srećković, Berber, & Filipović, 2022). They predict that ML will obviate the need for domain knowledge or expertise in science (Mayer-Schönberger & Cukier, 2013). The predictive success of ML is projected to replace the need for insight into causal mechanisms or the data-generating process (Mayer-Schönberger & Cukier, 2013; Spinney, 2022). Some of these statements echo proclamations that were once made of classical statistical method: that big data analytic tools promise to allow the raw data to "speak for themselves" (Levins & Lewontin, 1985).

Some of these claims of disruption fall under the heading of ML or AI *hype*—they issue from individuals swept up in a wave of drastically overselling the capabilities of presently existing ML techniques [1]. Others reflect what we term a *theory-free ideal* in science: aspirations of naïve empiricism rampant today within the special sciences. Subscribers to the theory-free ideal seek to purge science of what they see as epistemically compromising arbitrariness and subjectivity, which are brought on board when human critical thinking and conceptualisation of the phenomena under study play an essential role in shaping the empirical research programme or the conceptual tools wielded therein. Lastly, these claims have been motivated by a concern for the future of science. If the scientific process becomes automated, purged of theory, and overtaken by uninterpretable black-box algorithms with human domain experts pushed out of the loop, this third category worries that the epistemic products of science may cease to be accessible to human interpreters. Interestingly, whether motivated by optimism or pessimism for the future of science, assessments of the role of ML in science have converged upon the same essential thesis: science will undergo

[1] Often, though not always, because such individuals stand to materially benefit from this widespread cultural misperception

drastic change with the advent of ML-based methods, because such methods are theoretically unmoored or conceptually impoverished in a way that sets them fundamentally apart from existing instances of applied mathematics. We dub this second claim the *distinctness* claim.

If, indeed, the procedure of science or the status of knowledge produced in science are set to radically change, this merits serious engagement by scientists and philosophers of science. If, instead, as we will argue, this is baseless sensationalism—which has established a foothold in not only the public consciousness but in communities of relevant experts—this deceptive narrative ought to be challenged, for it will lead scientists and the public astray. The distinctness claim latches onto a real novelty in much ML deployed toward scientific ends: potential for misuse and lack of methodological standards. Instead of identifying this as the epistemic problem it represents, however, the distinctness claim functions to reify this (potential) misuse of ML-based tools into an account of how these tools normally function, how they necessarily function, or even how they normatively *ought* to function.

## 2 Disruption & Distinctness

Claims of disruption and distinctness do not originate from within philosophy of science—they have been repeated for some time now by scientists, engineers, and journalists. Such proclamations as they come from engineers and journalists, or made in passing by scientists providing reconstruction or meta-commentary on the subject of their own work, lack the detail and rigour that philosophers of science are held to in defending such claims, and for this reason we engage principally with claims of disruption and distinctness as they are expressed by philosophers of science in formal venues. It is our hope, however, that the argument generalises as a response to all instances of such claims, independent of disciplinary affiliation, interlocutor, or venue. We will look to three texts in the philosophy of science as representative of disruption and distinctness claims.

Boge (2022) speculates that a revolution in either scientific practice or its epistemic footing may be in store owing to the adoption of machine learning—specifically deep learning—methods. Boge's argument rests on the idea that deep learning is both instrumental in an idiosyncratic sense among modelling approaches in the sciences, and that it exhibits a novel kind of epistemic opacity to its deployers. These identifying facets of deep learning pose an impediment to understanding and explanation (in the scientific sense), especially when deployed in exploratory settings where the successful results of scientific enquiry will require novel concept formation. Owing to their divergence from standard mathematical modelling practices in the sciences, Boge claims, ML modelling techniques "have the potential to profoundly 'change the face of science'" (Boge, Grünke, & Hillerbrand, 2022, p.71).

Boge urges that the distinction between the procedure of classical mathematical modelling or computer simulation in science and the application of machine learning methods is that the former procedure begins with a conceptualisation

of the target phenomenon under investigation, while this step is absent in the use of ML. Especially in exploratory modelling contexts, the lack of background theory or conceptualisation of the target phenomenon is taken as an impediment to understanding. While Boge grants that DL models might represent, he holds that they fail to be explanatory for lack of theoretical context and conceptual content. Boge takes after de Regt in his stance on the relation between representational status and explanatory status: "for representational models to explain, they must also be constructed under the principles of an intelligible theory, where a theory is intelligible if it has certain qualities that 'provide conceptual tools for achieving understanding' (de Regt, 2017, p. 118)"(Boge, 2022, 54). Boge predicts profound changes to the practice and epistemic products of science because ML-based tools will fail to provide understanding or explanations due to their lack of theoretical or conceptual motivation and content.

In a similar vein, Sreckovic, Berber, and Filipovic (2022) differentiate machine learning techniques from standard practices in statistical modelling, arguing that statisticians employ theoretical assumptions, while machine learners do not (Srećković et al., 2022). Sreckovic, Berber, and Filipovic (2022) evaluate what they hold to be the key differences between traditional modelling approaches and machine learning methods in terms of the explanatory capacity of both and their capacity to elucidate causal relationships. Sreckovic et al diagnose the methods of machine learning as uninterpretable, and not resting on theoretical considerations. This, according to the authors, prevents the practice from getting at underlying causes and furnishing explanations of natural phenomena. The ability of ML techniques to provide prediction in the absence of explanation is projected by the authors to alter the landscape of how we conduct science.

"In contrast to explanatory-focused statistical models," Sreckovic et al argue, "ML models reach predictions without the theoretical backup that supplements the correlations found in the data with a potential causal interpretation" (Srećković et al., 2022, 160). Machine learning, they argue, is "theory-agnostic" in that "there are no a priori assumptions concerning the mechanism of the target phenomenon" (Srećković et al., 2022, 165). While the authors acknowledge a sort of disappearing line between ML and traditional statistical techniques, their emphasis is on drawing out broad characterisations of the two disciplines and what separates them. Whereas for "traditional statistics, standard models rely on the representation of underlying causal mechanisms, and they are used for retrospective testing of an already existing set of causal hypotheses...ML models are constructed based on data instead of theoretical assumptions about the target system. The purpose of such models is primarily forward-looking, i.e. to predict new observations" (Srećković et al., 2022, 166). Here, the contrast the authors draw between broadly "data-driven" and "theoretically-motivated" methods is telling. This distinction is not one the authors have introduced: such a divide between theory-driven or hypothesis-driven research and data-driven research is held widely among engineers and scientists. Sreckovic et al merely provision a philosophical exposition and justification thereof.

The majority of philosophers of science grappling with the existence and

scientific uptake of tools from ML have accepted the premises that machine learning is, in the first place, fundamentally different from existing approaches in applied mathematics and, in the second, that it will usher in sweeping changes to the epistemic products or practices of science. One noteworthy exception is a 2020 work by philosopher of science Mieke Boon. Boon (2020) argues against the thesis that machine learning methods will obviate the need for auxiliary or intermediary human conceptual apparatus in the generation of scientific knowledge. She argues that the reason that we grant any sort of a priori plausibility to statements to the effect that big data will usher in a scientific revolution flows from a shared implicit view of how science works—one which she argues to be in error. She labels this erroneous conception of science a "strict empiricism." Her goal is to "make plausible that on an empiricist epistemology the elimination of any human contribution to scientific knowledge is in fact already built in as a normative ideal...strict empiricist epistemologies indeed support the claim that objective, although opaque, data-models produced in machine learning processes can replace and may even be preferable to human-made scientific knowledge" (Boon, 2020, 46).

Boon advocates for the necessity of human capacities for conceptualisation, abstraction, and interpretation in every aspect of collecting, preparing, and manipulating data: "not only when setting up the data-generating instrumentation and seeing to its proper functioning, but also in assessing and interpreting the data, drawing relationships between data from different sources, and for making the distinction between 'real' phenomena and artifacts" (Boon, 2020, 59). Further, "[t]he necessity to prepare data that are about something in the real world also implies that phenomena are crucial in scientific practices, even when only aiming at the generation of data for machine-learning processes" (Boon, 2020, 57). As evidenced in these passages, Boon clearly takes data provenance and processing to be an interpretive affair. But for her argument against would-be empiricist dogma to work, she must take it axiomatically that data and data models are objective and worldly. This is part and parcel of the misconception of scientific process and products which we believe Boon seeks to argue against—the misconception which we are, in this paper, chiefly arguing against. Namely, the misconception of data as being raw, objective, and worldly—unmediated by human theorising and conceptual grasp on the target.

If we banish the idea that data is objective and worldly from the start, instead viewing data collection, cleaning, processing, and interpretation in an inference-licensing capacity as a fundamentally theory-mediated affair, Boon's contentions with empiricist epistemologies appear to dissipate. Perhaps the stumbling block is most easily seen in Boon's in-passing characterisation of the role of idealisation in mathematical representation. Boon claims that "machines are not confined by the kinds of idealizations and simplifications humans need to make in order to fit data into comprehensive mathematical formalisms" (Boon, 2020, 51). The idea that the role of idealisation in scientific representation ultimately serves the human-interpretability of our representations—and that idealisations are evitable or eliminable—is not, of course, novel or idiosyncratic to Boon. It is, however, revelatory of her commitments to the representational

properties of applied mathematics. Mathematical representation is conceptual work. Idealisation is essential to it. Use of ML-based tools in science thus cannot escape the necessity of idealisation.

Boon is a vocal proponent of a theory-laden conception of data. Yet her analysis of the prospects for machine learning in science appear to reveal inconsistencies in her view. Like Boge and Sreckovic et al., Boon concludes that applications of ML in science will fall short of providing understanding or explanation in virtue of being conceptually impoverished. This, on her view, sets applications of ML to scientific research intrinsically apart from "real science." "'[R]eal science' and machine learning technologies," she writes, "operate in very different domains and must not be regarded as competing" (Boon, 2020, 58). If data is necessarily theory-laden and conceptually-mediated, however, then it cannot be the case that ML-facilitated science is a theory-free or concept-free epistemic activity, because the use of ML in science will be necessarily inflected by the theoretical and conceptual commitments inherent to the data.

Boon, Boge, and Sreckovic et al each sign onto the idea that ML methods are theory-free, and hence distinct from canonical modelling methods in science. Boge and Sreckovic et al further contend that the widespread adoption of ML methods will catalyse disruptive change in science, while Boon argues that the theory-freeness of ML methods rules them out as viable tools for science. These scholars take the perceived differences between "normal science" or even "real science" and machine learning to amount to the degree to which they are theory-laden, theory-driven, or conceptually rich. As we will demonstrate in the subsequent sections, no use of ML in science is "theory-free," and those that aspire to this ideal tend to result in poor scientific practice.

# 3 Theory Ladenness

Even the most simplistic of experimental designs reveals the nature and extent to which data, and scientific practice at large, are "theory-laden." The very act of investigation involves commitment to the existence and in-principle measurability of some phenomenon and, if we are making measurements and performing quantitative analyses thereon, commitment to its quantitative nature. How we choose to measure a phenomenon generally includes a commitment to a kind of quantitative ontology of the phenomenon, e.g., is it categorical, ordinal, or cardinal? Measurement cannot be total, and therefore there is always a commitment as to what to look at experimentally and what to exclude. There is always a commitment to the appropriate level of abstraction at which to study the phenomenon in play in terms of such things as instrument settings like degree of magnification or periodicity of sampling. The very design of our instruments of measure and their calibration includes various commitments to the nature of the worldly phenomena under investigation. In fundamental physics, when we cool our instruments to reduce the contamination of our measurements by thermal noise, it is our prior theoretical grasp on the target phenomena, the physical systems under study, that motivates us to do so. "Data" is not physical

phenomena. "Data" is abstract representation of the results of direct observation or measurement which is capable of serving an evidential role in licensing inferences about physical phenomena.

In contemporary philosophy of science, the idea that data is theory-mediated is widely endorsed (Boyd & Bogen, 2009; Bogen & Woodward, 1988; Gitelman, 2013; Leonelli, 2019b). Philosophers of science have largely overcome "[t]he naïve fantasy that data have an immediate relation to phenomena of the world, that they are 'objective' in some strong, ontological sense of that term, that they are the facts of the world directly speaking to us" (Longino, 2020, 391) and accept now that there is "no pristine separation of model and data" (Lloyd, 2018, 397). Bogen (2016) argues that it is the very fact that data is not raw, that it is, in a sense, "impure" that makes it able to serve the meaningful epistemic role it does (Bogen, 2016). Boyd (Boyd, 2018; Boyd & Bogen, 2009) argues further that it is not in spite of, but owing to the theory-ladenness of data that empirical science garners us its epistemic results.

Philosophers of science now popularly profess allegiance to a theory-laden conception of data; certainly none nowadays outright defend an account of data as raw and objective. But the philosophers of science most attuned to how data is understood in the context of scientific practice note that in many philosophical accounts, data is still *treated* as raw and objective, "as reliable information source—a mere "input" into processes of modelling" (Leonelli, 2019b, 4). Leonelli (2018) argues that mainstream accounts from within philosophy of science—though they might profess otherwise—tend to treat data as representational, and understand this representation relation in terms of the recapitulation of worldly structure in data. "Philosophers tend to assume that data have some sort of representational content, in the sense of instantiating some of the properties of a given target of investigation in ways that are mind-independent" (Leonelli, 2019b, 4).

Although no philosophers of science today would appear to explicitly endorse an interpretation of data as objective and worldly, unfortunate relics of this view remain widespread, often in the form of a conception of data as mere "empirical input for modelling" hence "implicitly accepting a view of data as intrinsically reliable representations of the world" (Leonelli, 2019b, 4). Leonelli (2018) investigates "the different extents to which theory—understood broadly as a set of theoretical commitments and goals—impinges on inferential processes from data" (Leonelli, 2019b, 22). In several book-length treatments of the use and interpretation of data in scientific practice (e.g., (Leonelli, 2018, 2019a; Leonelli & Tempini, 2020; Leonelli & Beaulieu, 2021)), Leonelli concludes that there is no place in scientific practice in which we have data that is not already, to some degree, shaped by our existing conceptual or theoretical grasp on the phenomenon, commitments to epistemic goals and questions to be answered, idealisations, and auxiliary assumptions.

# 4 Machine Learning and Deep Learning

## 4.1 Machine Learning

ML is fundamentally a set of mathematical and computational tools for drawing inferences (learning) from data. As a discipline, it has historically been divided into three broad categories: supervised learning, unsupervised learning, and reinforcement learning—only the first two of which will be relevant for our purposes in this essay. Supervised learning involves training a model to predict an unknown property of an observation from a known property, e.g., in object recognition, inferring the presence or absence of an object (unknown property) in an image from the pixel values (known property). The trained model approximates a function, and the function it has learned is (we hope) able to extrapolate to correctly fit or categorise unseen instances—instances which lie outside of the training dataset. We set aside subsets of labelled data in holdout or validation sets to check that we have not, for instance, overfit to uninformative idiosyncracies of the training set.

For our purposes, what is salient is that supervised learning always essentially seeks to learn the rule expressed in the training set, which human users or developers have deemed to be the learning objective for their epistemic purposes. The task is to generalise this rule, which relies on the premise that the data on which the model is deployed will be sufficiently similar to the training data. Supervised learning is always, therefore, a straightforwardly theory-laden exercise. The data on which supervised models are trained reflect human determinations of salience, of what the meaningful patterns to be distilled are, of what are outliers to be ignored, whether the data sufficiently resembles future data to be collected, and ultimately what the criteria are for successful prediction/classification. Supervised learning cannot be free of human conceptual influence because what we are attempting to extract from data is precisely the human concepts the data is imbued with.

Unsupervised learning is a category defined primarily in opposition to supervised learning. It encompasses dimensionality reduction and embedding, generative modeling, and clustering, to name only a few of the disparate practices which fall under this umbrella. The unifying feature of unsupervised methods is that they are tasked with extracting the contours of data; the data on which the models are trained do not embody a preexisting conceptualisation of the "right answer" that the models should learn. For instance, we might employ an algorithm to partition a dataset into some $k$ number of groupings such that the groupings minimise the within-group distance between datapoints along every featural dimension expressed in the dataset—this technique is known as *k-means clustering*. At a glance, the task of unsupervised learning appears not to be shaped by preexisting scientific theory or human conceptualisation of the phenomena under study in the same way that supervised learning is; whereas SL is always an attempt to replicate an existing decision rule, UL attempts to extract meaningful patterns, trends, or clusters from the "raw data" alone. From this alone, we might reason that unsupervised learning methods

are somehow divorced from theorising, making it the kind of ML that will "alter the landscape of science," as many have claimed. Such a conclusion, however, is only reached if we are ignorant as to the theory-ladenness of data and model evaluation, and the constrained application of unsupervised learning techniques in science. Unsupervised learning is typically relegated to the role of fulfilling relatively minor tasks within a more involved scientific pipeline—data preprocessing, curve-fitting, dimensionality reduction, and whatnot. It is not used to blaze new trails of scientific discovery. Attempts to saddle unsupervised learning techniques with a more robust role in scientific practice under the assumption that the methods are "theory-free" or "obviate the need for theory" lead to pseudoscientific practices, as we illustrate in a case study in section 5.2.

## 4.2 Deep Learning

ML refers properly to a broad class of computational methods, and the use cases for ML methods in science are consequently very broad. The disruption and distinctness claims we take as the target of this essay appear to centre around deep learning methods. It is worth noting here that machine learning properly denotes much, much more than mere deep learning (DL), and that various machine learning methods are already widely deployed across the sciences—as they have been for some time. Even restricting our analysis to the methods of deep learning alone, however, we still find both the methods and the use cases to be quite diverse. Potential applications of DL techniques in scientific enquiry range from anomaly discovery in cosmological surveys and sifting signal from noise in particle collision events, to nanomaterials discovery, biomaterials discovery, drug discovery, and protein folding, to serving as a model of image processing in the primate visual system. Generative Adversarial Networks (GANs) alone, for instance, can be useful in materials discovery or drug discovery, in preprocessing data or images for later automated classification, or in "filling in the blanks" of discrete time step images—for instance, neuroimaging or cell development—to create a continuous time evolution. We can utilise the methods of DL to approximate solutions to stochastic PDEs. We can use DL to probe the latent space of biomaterials for as-yet-unimagined cold-tolerant hydrophilic protein structures. We can use DL to represent how the mammalian visual cortex engages in object recognition. We can use DL to draw a line between particle collision events likely to be interesting and those which are uninformative. There remain such a plurality of use cases for deep learning in science that any two have little more to do with one another than any two arbitrary instances of applied mathematics. The techniques of machine learning can be utilised in science for problems which are as basic as linear regression or image pre-processing. The more impressive results, however—the results commanding the attention of the broader scientific and philosophical community—are far more sophisticated than these, and all involve modern deep learning architectures.

Sensationalism about the capabilities of deep learning models abounds today, and it is partially the aim of this text to combat it. With that proviso, there is much that DNNs are capable of which is of considerable philosophical,

mathematical, and scientific interest, and which marks an abrupt (and, according to many, inexplicable) break from the capabilities of traditional statistical methods and more mundane ML modelling tools. The ability of these models to generalise as well as they do remains unexplained within existing theoretical frameworks (e.g., those that provide generalisatuon guarantees for supervised learning). The novelty of DL methods rests essentially on 1. their ability to function in very high-dimensional regimes (e.g., computer vision and natural language) and 2. their ability to do so without the need for painstaking feature extraction/feature engineering(which dominated pre-DL methods. This latter feature of DL is, at least partially, responsible for the claim that DL "obviates the need for domain expertise." Features refer to dimensions of a dataset. If each datapoint represents an individual, the features might represent height, weight, smoker status, and blood pressure. When we refer to *big data* or *high dimensional* datasets, we are referring to datasets with potentially billions of datapoints and hundreds of thousands of dimensions. Feature engineering or extraction refers to conceptual and statistical means of organising datasets according to the featural dimensions they express. DL methods are generally capable of extracting meaningful statistical relationships without necessarily relying on hand-engineered features. However, as we shall see in a detailed case study in the following section, this only means that theoretical considerations come in elsewhere in the model development pipeline.

# 5 Theory-Ladenness in ML

## 5.1 The Unreasonable Efficacy of AlphaFold

Far and away the most impressive result that ML methods have achieved for science is AlphaFold 2.0. To appreciate the unprecedentedness of the AlphaFold results, we must first appreciate the scientific problem it is confronted with. The problem of protein folding is notoriously difficult. There is very little that we can say from the genotypic specification of a particular protein about how it will fold. Mapping from sequences of adenines, cytosine, guanines, and thymines to a menagerie of amino acids is straightforward, as is predicting the polypeptide chains these amino acid sequences will form. What mess of three-dimensional spaghetti those amino acid chains will assume once synthesised, however, is another matter entirely. This is an essential problem for the biomedical sciences. The three-dimensional anatomy of protein structure is determinative of its function and is thus a crucial object of scientific inference.

To truly comprehend the difficulty of the protein folding problem—and how the methods of machine learning were able to get around it—we have to recognise that protein structure is understood at four levels. DNA is a string composed of four alternative base pairs. It encodes information in sequence. When proteins are assembled, that DNA is read, codon by codon, and a polypeptide chain is built up from twenty amino acids on the basis of these instructions. These amino acid sequences are dubbed the "primary structure" of a protein.

All amino acids are composed of the same base molecular structure of 9 atoms, which will bond together to form the backbone of the polypeptide chain. From this molecular backbone extends the R-group or side chain, the determinant of the amino acid's "flavour." The secondary structure of a protein refers to the morphology that polypeptide chains take on on their own, owing to bonding patterns in the backbone. The morphology of these peptide chains results from local interactions between adjacent and semi-adjacent molecules in the backbone of the peptide chain. Owing to the periodicity of the placement of amino acids with certain valences (and other molecular-bond determining features) in the chain, they will typically either form what are known as $\alpha$ helices or $\beta$ sheets. Up until this point things have remained relatively straightforward: we have a basic, repeated molecular structure and its self-interaction in the form of hydrogen bonding.

The tertiary structure of a protein is determined by the $R$-groups of the amino acids. Recall that these come in twenty flavours. Recall that virtually all forms of non-covalent bonding are available to these molecules now. Recall that amino acids can exhibit hydrophobic and hydrophilic proclivities. If a protein is composed of more than one polypeptide chain, it will have a quaternary structure as well. At the tertiary and quaternary levels of protein structure, we have advanced from assembling text from bit strings to attempting to predict all of the ways in which several distinct kinds of spaghetti thrown together in a pot can cohabitate, given six dimensions along which spaghetti substructures may or may not like to interact.

At first blush, this seems like an unsolvable problem. The initial trick—the trick that gets existing bioinformatic solutions off the ground—lies in noting that when we have a variant in one amino-acid we can see what *non-local* variants tend to co-vary along with it. This begins to tell us something about what might be touching what in the tertiary and quaternary protein structures. Still a difficult problem, but more manageable. These associations of covarying amino-acid substitutions lend us what is known as a *protein contact map* which further lends us a *multiple sequence alignment* (MSA).

The AlphaFold team created their own database of protein structures—now the largest existing database of its kind—by scraping[2] existing publicly-available databases. DeepMind's AlphaFold 2.0 runs queries on an amino acid sequence in its pre-processing stage to obtain a multiple sequence alignment (MSA). Any modern approach to predicting protein structure begins with an amino acid sequence as input. As we have noted, given the state of modern biological knowledge, it is trivial to determine amino acid sequences given the protein's genetic blueprint. To construct the inputs, AlphaFold queries protein structure databases to assemble an MSA. In addition to the primary amino acid sequence and MSA, AlphaFold was also supplied as input database-derived *templates*—three-dimensional atomic maps—for a small number of sufficiently similar homologous protein structures. The templates and the MSA are rendered together to create what the AlphaFold team dubs a *pair representation*.

[2] I.e., automatically extracting web data.

AlphaFold treats the prediction of 3-dimensional protein structure from these pair representations and MSAs as a graphical problem, rendering the representations in the primary trunk of the model architecture into gradated bitmaps. The problem formulation for the Deepmind team was to "view the prediction of protein structures as a graph inference problem in 3D space in which the edges of the graph are defined by residues in proximity" (Jumper et al., 2021, 585). The core structure of AlphaFold 2.0 is a transformer—a form of DNN architecture which is easier to train and outperforms competing architectures by parallelising and better attending to higher-level contextual factors in the training data. AlphaFold passes both the MSA and the pair representation back and forth through the trunk of the model for a set number of iterations (48 blocks), progressively refining the representations, and allowing the two distinct representations (MSA and pair representation) to influence one another as each is refined. The output of this refinement procedure is then, in the final stage, fed to a generative neural network which produces a plausible candidate 3-D protein structure. The 3D protein structure is then passed, with MSA and pair-representations, back through the trunk. This is repeated for three iterations until a final predicted 3D protein structure is achieved.

Let us draw out what is salient about this scientific procedure for our analysis. Our aim is to show that theoretical considerations are playing an essential role at the stage of data provenance and engineering, the stage of architecture design, hyperparameter selection and model training, and at the stage of model evaluation and interpretation.

Theory integration comes in at the level of the data in terms of what the data ultimately represents and how it is imbued with that representational content. Taking on board the notion of theory-laden measurement, we understand that the data on which AlphaFold is trained is richly structured by existing empirical knowledge of the target domain and our theoretical understanding thereof. AlphaFold sits atop a wealth of domain knowledge about the form and function of proteins. Theory also comes into play in how the data is handled for the specific task in question and how it is made to serve as evidence in this task. AlphaFold is, at its core, an instance of (semi-)supervised learning. The exercise is premised on the idea that the rules of association between amino acid sequences and three dimensional protein structure lie latent in cross-taxa protein structure data. It is further premised on the supposition that the systematic breakdown in protein structure and function resultant from certain amino acid substitutions can be leveraged to learn the complex bonding affinities governing 3-dimensional protein structure. Part of what is noteworthy in this case study is the insight to take the publicly available data and turn it into novel representational forms in multiple places: combining MSAs and templates to create pair representations, and projecting those into effective heatmaps of sequence-structure associations so that the inference task could be treated like a graphical problem.

The architecting of the various model components utilised in AlphaFold 2.0 was similarly bound to theoretical considerations. AlphaFold is not a domain-generic model; the model architecture is hand-tailored to the specific task of learning to predict three dimensional protein structure from MSAs and pair

representations—a novel representational form for the task. AlphaFold 2.0 employs a transformer network that is designed to iteratively refine progressively more accurate guesses at the true protein structure. The transformer trunk utilised in AlphaFold was created to combine and refine representations of the specific form it is fed in a novel training and deployment procedure. Perhaps the most strikingly theory-laden aspect of AlphaFold 2.0 is the engineering of specially tailored loss functions. In training a DNN, a loss function governs how the distance metric is calculated between present output and desired output of the model (in a typical neural network training regime, the error is then back-propagated through the network to update the model's parameters). In specifying the loss function, machine learners are able to express precisely what it is that they are interested in learning for a particular task. In AlphaFold 2.0, the loss function is heavily tailored to the problem of predicting folded protein structure from amino acid sequences. The researchers employed "a loss term that places substantial weight on the orientational correctness of the residues" (Jumper et al., 2021, 585). Loss terms specific to the learning of various structural features of protein folding along a number of dimensions were employed at all stages of training and fine-tuning: "satisfaction of the peptide bond geometry is encouraged during fine-tuning by a violation loss term" (Jumper et al., 2021, 586-587).

Finally, model-evaluation, that is, judging the success of the trained model and interpreting its results requires integrating the resulting predictions of AlphaFold into existing biological knowledge. We can only judge the success of such a model when it is understood against the backdrop of our prevailing scientific accounts. We can likewise only put the results of such a modelling effort to *use* when we have accomodated them within a theoretical framework.

## 5.2 Transcriptomics

Single-cell transcriptomics is a method for inferring cellular-level gene expression. The technique is utilised for identifying cell populations, modelling transcription dynamics, inferring the developmental trajectories of cellular populations, and monitoring changes in cell populations relative to health status. Single-cell transcriptomics emerged with the availability of massive quantities of high throughput RNA sequencing and expression data. It is typical in such exercises to be working with datasets which possess hundreds of thousands of feature dimensions; for this reason, researchers typically employ dimensionality reduction techniques. Dimensionality reduction is a (unsupervised ML) method of mapping a high-dimensional dataset to a lower-dimensional space—or *embedding* higher-dimensional data in a lower-dimensional space. Dimensionality reduction techniques are used to distill essential patterns from large datasets, make analyses tractable, and isolate signal from noise.

A now well-established workflow in single-cell transcriptomics involves applying dimensionality reduction techniques sequentially to high-throughput RNA expression data; first linear methods which reduce the dataset to tens of dimensions using principle component analysis (PCA) or analogous techniques of

dimensionality reduction, followed by one of two purpose-built two-dimensional nonlinear reductions: UMAP or t-SNE. The method produces visualisations for exploratory data analysis. A scientist cannot very well eyeball a 250,000-dimensional manifold and distill from it useful and meaningful information (or eyeball it at all). A two-dimensional embedding, however, might very well reveal visually intuitive information about cell populations and trajectories. However, as Chari and Pachter (2021) demonstrate, this now standardised procedure in single-cell transcriptomics lacks theoretical motivation, represents poor statistical practice, and is effectively incapable of providing meaningful biological information; instead it creates the opportunity for erroneous interpretation (Chari & Pachter, 2021).

In a series of analyses, the researchers demonstrated that the practice of repeated application of dimensionality reduction techniques introduced heavy distortions and was incapable of preserving the interpretively salient features of the datasets under investigation: local and global structure, distance, and continuousness (Chari & Pachter, 2021). Interpretive practices surrounding the resultant visualisations, they concluded, led to erroneous or conflicting conclusions. Chari and Pachter (2021) found that the combined use of supervised and unsupervised ML methods in single-cell transcriptomics was haphazard:

> "[T]he same k-nearest neighbor (knn) graph constructed from the higher dimensional PCA space is passed to both the clustering algorithm and the embedding algorithm...the embedding is then not an independent assessment of clustering results and is likely to form clusters that match the knn graph even if that graph does not represent the 'original' underlying manifold. Together, the use of such embeddings to imply or infer continuous relationships then becomes an arbitrary endeavour, with a user unable to trust seemingly dramatic connections or isolated populations, and likely to choose what seems most appealing" (Chari & Pachter, 2021, 14).

In one particularly striking example, Chari and Pachter projected transcriptomics datasets onto arbitrary shapes (a flower, von Neumann's elephant) and found that they preserved the interpretively salient features—local and global structure, distance, and continuousness—commensurate with, or better than, the resultant embeddings from PCA $\rightarrow$ t-SNE or PCA $\rightarrow$ UMAP workflows (Chari & Pachter, 2021).

These techniques fall under the heading of what we term *Rorschach research methods* or *intuition laundering*: interpreters of the results of these data analysis methods are free to cast upon them whatever intuitive interpretation appeals to them, wielding the graphics to lend supposed empirical support to their claims. Such pseudoscientific practices are likely to emerge anywhere the methods of ML are employed without adequate theoretical grounding and statistical literacy. Chari and Pachter (2021) demonstrate that semi-supervised learning methods and targeted embeddings for specific featural dimensions are capable of illucidating far more than the naïve methods they critique. Such approaches,

however, require domain expertise, critical thinking, and being able to both identify and (statistically) articulate what you are looking for—characteristics markedly absent from the t-SNE/UMAP workflows under scrutiny.

### 5.3 Takeaways about ML in Scientific Practice

AlphaFold is a case of resounding success; perhaps the greatest win for ML in science to date. No other application of ML to science has achieved quite so stark an advantage over pre-existing techniques. Many applications of the tools of ML to science, by contrast, have been run of the mill: automating laborious processes, achieving minor gains in efficiency or accuracy over human classification or "analogue" statistical techniques without notable breakthroughs in what sort of knowledge could be gained by their use. Many scientists have also faced great frustrations in incorporating computational tools into their research paradigms, either because they were attempting to utilise ML in an untenable, "theory-free" manner or because they faced difficulty in their attempts to imbue ML-based tools with the requisite theory or domain knowledge.

Researchers in the biomedical sciences bemoan the fields' recent infatuation with the tools of ML in operation with its longstanding "theory-aversion" (what we term a "theory-free ideal" in science, in analogy to Douglas' value-free ideal (Douglas, 2009)) (Coveney, Dougherty, & Highfield, 2016). Incorporating theoretical principles into ML-assisted and big data-fueled research can prove difficult, and is unlikely to happen when institutional and publishing incentives overwhelmingly favour the collection of higher volumes of data and the adoption of novel computational tools over critical thinking and principled research design. In fundamental physics, by contrast, the need for theoretically-informed models is more apparent and is met with less resistance. Karniadakis et al review methods of incorporating physical principles into applications of DL in physics (Karniadakis et al., 2021). Incorporating theory into ML-assisted scientific practice is no simple matter, but work of this kind reveals both its possibility and its necessity.

The use of unsupervised learning in physics is now concentrated on "physics-informed" architectures—forcing the model to conform to the form of a known e.g., physical, principle. These methods, unlike "naïve" unsupervised cluster or regression techniques, which can only occupy relatively simplistic intermediary calculational or bookkeeping roles, can play a far more central role in research. This is precisely *because* they have the conceptual resources to serve a meaningful role in empirical research. In this sense, the conclusion we reach aligns with some of the reasoning in Boge, Sreckovic et al., and Boon: theory-involvement is a requisite feature of our conceptual instruments in science for them to be able to elucidate previously unknown features of our natural world from data. The issue is that the perfectly theory-free vision of ML (or DL, or unsupervised DL) in science that is the target of these scholars' critiques either singles out a strawman or a failure case.

# 6 Conclusion

It is widely believed of the methods of ML—as reflected in the texts from philosophy of science which we have analysed in this paper—that if loosed on enough data they are capable of discovering meaningful patterns, natural joints, or mind-independent truths of their own accord, without any intervention by human theorising or conceptualisation of the target system. There is a naïve version of this view that only those unfamiliar with the ins and outs of applying ML methods could hold onto. However, a more sophisticated version of the thesis also exists and is commonly held even by engineers, researchers, and practitioners building and deploying ML-based tools. This is the idea that unsupervised learning tools are capable of discovering mind-independent natural patterns or boundaries in a "principled" manner without the need for any arbitrarity or human input. If we believe this, and if we also believe the techniques of ML to be "opaque" or "uninterpretable" in some novel way—in a way that sets them fundamentally apart from existing conceptual tools or instruments in science—then what is learned via these tools will be inscrutable to human scientists. If these tools are then adopted widely in scientific practice, this will then entail radical change to the varieties of epistemic outputs science is capable of generating. As we have argued in this text, the ideal of theory-free learning via ML from "raw data" is a confused one.

This conclusion is overdetermined when we consider accounts from philosophy of science and theoretical computer science of constraints on inductive generalisation. Inductive inference is the procedure of gaining knowledge by extrapolating from a limited number of instances to a more general class—the fundamental task of ML. According to Norton (2003), an account which he dubs the *material theory of induction*, successful inductive inference is never licensed by universal, domain-generic formal rules, but always proceeds by the application of local rules warranted by hard-won empirical (material) facts tied to a specific line of research (Norton, 2003). There is no one inductively valid formula to rule them all. The way in which the methods of ML and their epistemic aims are portrayed by those buying into and perpetuating ML "hype" is as though they seek to (and, indeed, will eventually) discover such a universally valid formal principle of inductive inference. Such a project is a doomed one. Learning theory has itself independently discovered the impossibility of a universally valid domain-generic inference rule: the no free lunch theorems (Wolpert & Macready, 1997). While these results obtain only in a very artificial setting, the moral they deliver is an important one for ML in practice: inductive inference only works in virtue of having learned domain-specific inductive biases. Claims of disruption and distinctness regarding the role of ML in scientific practice that are premised on a conception of applied ML as free from theoretical considerations and capable of generalising without inductive bias or the importation of domain knowledge thereby fail.

Claims to the effect that ML/DL in science is free from theoretical considerations are, we hope to have shown in the course of this paper, misguided. The idea that ML will "disrupt," "revolutionise," or "change the face of" scientific

methods or outputs are fueled by sensationalism and ignorance to the nuance of ML deployed in scientific practice. In reality, there is a disappearing line between tried and true methods of mathematical modelling or statistical analysis in science and ML-based methods. In this essay, we have largely focussed on illustrating some of the ways that ML is utilised in and in support of empirical research. The aim has been to defeat something of a cartoon vision of ML in science which we take to be the actual target of distinctness and disruption claims—for the claims fail to make sense when addressed at actually existing or possible deployments of ML/DL methods in science. But this confusion surrounding ML in science has another likely source: namely, a cartoon vision of the scientific method itself.

There exists a great deal of methodological diversity in scientific practice today, as there have been historically, and a great deal of diversity in the ways in which theory plays a role in such practices. Yet the practical and stated aims of philosophy of science in the twentieth century, its *raison d'être*, lay in characterising *empirical knowledge* and *the scientific method*. The public shares, if not in the details of such a vision, in the unitary conception of knowledge and empirical method. It is what we are taught from an early age from our schoolbooks and what gets echoed to us throughout our lives from popular media and expert scholarship alike. In stark contrast to its role throughout much of the previous century, philosophy of science today has embraced *pluralism*. Pluralism refers to the idea that there is not one viable scientific method that holds across disciplines, subject matter, epistemic aims, and stages of scientific maturity, but a variety of them. Many proponents of such a view connect it to the associated idea that we are better able to understand natural phenomena by coming at them simultaneously from a number of methodological angles and triangulating.

The central problem inherent to pluralism as practiced by the majority of contemporary philosophers of science is as follows. They take their work to consist in studying the methods of particular, highly-constrained areas of science in great detail and to provide rational reconstructions of what the work entails and how the results are achieved, with almost complete deference to the scientist, to the professed or institutionally-acknowledged validity of their methods, and to the scientist's own introspective accounts of what they are doing. This presents a major issue if we take the key epistemic function of philosophy of science to rest in providing the means to differentiate between epistemically sound and epistemically unsound scientific practices. Scientists are, on the whole, powerfully disincentivised to proclaim their own work scientifically unsound.

It is our belief that the public—and many philosophers of science, implicitly—have held onto a monolithic vision of scientific methods and epistemic outputs because the pluralistic approach fails to deliver on this front. The accounts of ML in science we take here as the targets of our critique get things wrong precisely in the failure to acknowledge a distinction between sound and unsound applications of an epistemic technology. Once we are willing to draw such lines in the sand, we reveal that most of the assessments of ML from philosophers of science amount to the observation that we might do shoddy science with

ML—or that ML even increases the likelihood of doing shoddy science. The reality is that we are already doing shoddy science—science that is shoddy in precisely the same ways (if marginally slower) as shoddy science might be done as aided by the tools of ML. We are already rearing generations upon generations of scientists to solve nature's puzzles by "throwing math at it" without teaching them to speak the language of math (or, for that matter, to speak the language of nature). The mythos of "theory-freeness," meanwhile, merely feeds into cartoon characterisations and popular narratives surrounding science and ML which support such malpractice. Philosophers of science, we argue, have a duty to disabuse scientists and the public of such misapprehensions and promote a more nuanced understanding of ML and its reliance on theory in their place.

There is a final misconception responsible for the understanding of ML as a theory-free endeavor. It is a misconception about the epistemology of applied mathematics, which we believe to have originated from folk intuition but which philosophers of science have only served to cement and to legitimise. Most philosophers of science defend, at least implicitly, the idea that there are two quite distinct varieties of applied mathematics. There are, in the first place, formal models which represent phenomena in nature; which stand in for natural systems, allowing surrogative reasoning about them. In the second place, there are tools from mathematics which do not represent contingent truths about nature, which serve mere calculational or bookkeeping roles. This dichotomisation of applied mathematics echoes how working scientists, modellers, and applied mathematicians *conceptualise* of their own work but not the actual *practice* of wielding applied mathematics. As we have seen illustrated in the case studies of ML tools used in science in 5.1 and 5.2, even instances of applied mathematics described as "mere data modelling" or "mere curve-fitting" involve representation of the target phenomena, and their epistemic success depends on engaging in explicit reasoning about the representational dimension of the work. No applied mathematics is agnostic to what it is being applied to. As the case of stacked dimensionality reduction procedures in single-cell transcriptomics demonstrates, failure to consider the appropriateness of a mathematical tool to the system under study and the epistemic task at hand results in futile scientific activities.

Advancing the state of the discourse away from false dichotomies and misdirected concerns is essential, for there is much that is interesting and potentially novel about ML/DL and its potential use cases in science. Where to localise theoretical considerations in DL-based scientific workflows differs, along various dimensions, from a certain canonical mode of doing science. On a received, roughly hypothetico-deductive view of experimental science and statistical modelling, we are typically formulating hypotheses and going out to collect data capable of adjudicating between our hypotheses. Thus the ways in which our conceptual grasp on the target phenomena come into play in how the data represents the target are specific to the epistemic concerns of a particular scientific/modelling exercise.[3] In big data analysis and applied ML, we are often

[3] This in stark contrast to Bogen and Woodward's (1988) claim that data are limited to

handed data corpora or else construct them from amalgamations of preexisting datasets. This means that a significant amount of the interpretive work, the work of mapping the data onto target phenomena—imbuing it with representational status and content—is work done before we are ever in contact with the data. Theoretical or interpretive work typically comes in again in the problem formulation, in the engineering of a model architecture and specification of loss, and in training. Theoretical considerations further come in at the level of model evaluation, in our formal assessments of the success of the exercise. Finally, such considerations come into play in what we take ourselves to have learned from the model output and, effectively, in *how the model is wielded*. The nature and role of theoretical considerations in such a scientific workflow are somewhat baffling if our template for how science works is based in, say, classical mechanics.

The landscape of science is also undergoing significant changes today that are worthy of popular and philosophical scrutiny. There are many very real and substantive changes to scientific practice occurring nowadays, to the social, institutional, state, and economic infrastructures that support it, and to the knowledge economies it results in. These include the fragmentation and specialisation of science, the proceduralisation of science, its automation, the progressive increase in the distribution of intellectual labour it involves, the extraction of the knowledge of domain experts and its mechanisation and codification into formulae. Reactions to the adoption of ML in science have largely framed ML as catalyst to these changes. We wish to counter that we can instead view ML as symptomatic of a much older and deeper trend in the development of scientific practice, one which often replicates the form of the society in which scientific practice is embedded in its social structure, its economic model, and its governance. The causal arrow runs at least as much from the automation of scientific practice to the adoption of the tools of ML in science as it does in the reverse.

To make sense of the present day landscape of science and the directions in which it is evolving, we will require a philosophy of science of machine learning. This must, however, be a philosophy of science willing to cast off an outdated, monolithic, and overly-restrictive conception of scientific methods and the epistemic outputs of science. This must be a philosophy of science willing to weigh in on debates and draw boundaries between admissible practices and the pseudo-scientific or pseudo-statistical. It must be a philosophy of science that is not conned by the hyperbolic narrative of the inscrutability of machine learning methods; that is willing and able to comprehend the techniques and how they are wielded to empirical ends.

# 7 Acknowledgements

The author is indebted to the following scholars for detailed feedback, critical engagement, and illuminating discussion without which this manuscript would

serving an evidentiary role in a particular experimental context (Bogen & Woodward, 1988).

never have come to fruition: Colin Allen, Robert Batterman, Eric and Rebecca Bruning, Daniel Burnston, David Childers, Nic Fishman, Leif Hancox-Li, Zachary C. Lipton, Arya McCarthy, Amira Moeding, John D. Norton, Angela Potochnik, Jennifer L. Whyte, and Tan Zhi Xuan.

# References

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired magazine*, *16*(7), 16–07.

Boge, F. J. (2022). Two dimensions of opacity and the deep learning predicament. *Minds and Machines*, *32*(1), 43–75.

Boge, F. J., Grünke, P., & Hillerbrand, R. (2022). *Minds and machines special issue: Machine learning: Prediction without explanation?* Springer.

Bogen, J. (2016). *Empiricism and after*. Oxford University Press.

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *The philosophical review*, *97*(3), 303–352.

Boon, M. (2020). How scientists are brought back into science—the error of empiricism. *A Critical Reflection on Automated Science: Will Science Remain Human?*, 43–65.

Boyd, N. M. (2018). Evidence enriched. *Philosophy of Science*, *85*(3), 403–421.

Boyd, N. M., & Bogen, J. (2009). Theory and observation in science. *Stanford Encyclopedia of Philosophy*.

Chari, T., & Pachter, L. (2021). The specious art of single-cell genomics. *BioRxiv*, 2021–08.

Coveney, P. V., Dougherty, E. R., & Highfield, R. R. (2016). Big data need big theory too. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2080), 20160153.

Douglas, H. (2009). *Science, policy, and the value-free ideal*. University of Pittsburgh Pre.

Duarte, J., Han, S., Harris, P., Jindariani, S., Kreinar, E., Kreis, B., ... others (2018). Fast inference of deep neural networks in fpgas for particle physics. *Journal of Instrumentation*, *13*(07), P07027.

Gitelman, L. (2013). *Raw data is an oxymoron*. MIT press.

Hey, A. J., Tansley, S., Tolle, K. M., et al. (2009). *The fourth paradigm: data-intensive scientific discovery* (Vol. 1). Microsoft research Redmond, WA.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... others (2021). Highly accurate protein structure prediction with alphafold. *Nature*, *596*(7873), 583–589.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*(6), 422–440.

Leonelli, S. (2018). La ricerca scientifica nell'era dei big data.

Leonelli, S. (2019a). *Data-centric biology: A philosophical study*. University of Chicago Press.

Leonelli, S. (2019b). What distinguishes data from models? *European journal for philosophy of science*, *9*(2), 22.

Leonelli, S., & Beaulieu, A. (2021). Data and society: A critical introduction. *Data and Society*, 1–100.

Leonelli, S., & Tempini, N. (2020). *Data journeys in the sciences*. Springer Nature.

Levins, R., & Lewontin, R. (1985). *The dialectical biologist*. Harvard University Press.

Lloyd, E. A. (2018). The role of "complex" empiricism in the debates about satellite data and climate models. *Climate modelling: Philosophical and conceptual issues*, 137–173.

Longino, H. E. (2020). Afterword: Data in transit. *Data journeys in the sciences*, 391–399.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Norton, J. D. (2003). A material theory of induction. *Philosophy of Science*, *70*(4), 647–670.

Rajaraman, A., & Norvig, P. (1998). Virtual database technology: transforming the internet into a database. *IEEE Internet Computing*, *2*(4), 55–58.

Spinney, L. (2022). Are we witnessing the dawn of post-theory science. *The Guardian*, *9*, 2022.

Srećković, S., Berber, A., & Filipović, N. (2022). The automated laplacean demon: How ml challenges our views on prediction and explanation. *Minds and Machines*, *32*(1), 159–183.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, *1*(1), 67–82.