The epistemological foundations of data science: a critical analysis

Jules Desai[1], David Watson[2], Vincent Wang[3], Mariarosaria Taddeo[4,5], Luciano Floridi[4,6*]

[1] Faculty of Philosophy, University of Oxford, Radcliffe Humanities, Woodstock Road, Oxford OX2 6GG, UK

[2]Department of Statistical Science, University College London, 1-19 Torrington Place, London WC1E 7HB, UK

[3]Department of Computer Science, University of Oxford, Parks Road OX1 3QD, UK

[4]Oxford Internet Institute, University of Oxford, 1 St Giles', Oxford, OX1 3JS, UK

[5]The Alan Turing Institute, British Library, 96 Euston Rd, London NW1 2DB, UK

[6]Department of Legal Studies, University of Bologna, Via Zamboni, 27, 40126 Bologna, IT

*Email of correspondence author: luciano.floridi@oii.ox.ac.uk

Abstract

The modern abundance and prominence of data has led to the development of "data science" as a new field of enquiry, along with a body of epistemological reflections upon its foundations, methods, and consequences. This article provides a systematic analysis and critical review of significant open problems and debates in the epistemology of data science. We propose a partition of the epistemology of data science into the following five domains: (i) the constitution of data science; (ii) the kind of enquiry that it identifies; (iii) the kinds of knowledge that data science generates; (iv) the nature and epistemological significance of "black box" problems; and (v) the relationship between data science and the philosophy of science more generally.

Keywords

Black box; Data-driven science; Data science; Epistemology; Foundationalism; Philosophy of science.

1. Introduction

Data science has become a mature field of enquiry only recently, propelled by the proliferation of data and computing infrastructure. While many have written about the philosophical problems in data science, such problems are rarely unified into a holistic "epistemology of data science" (we avoid the more generic expression "philosophy of data science" – more on this presently). In its current state, this epistemology is vibrant but chaotic. For this reason, in this article, we review the relevant literature to provide a unified perspective of the discipline and its gaps; assess the state of the debate; offer a contextual analysis of the significance, relevance, and value of various topics; and identify neglected or underexplored areas of philosophical interest. We do not discuss data science's GELSI (governance, ethical, legal, and social implications). They already receive considerable attention, and their analysis would lie beyond the scope of the present work, even if, ultimately, we shall point to obvious connections. It seems clear that data science's epistemology and ethics (in the inclusive sense of GELSI indicated above) may need to find a unified framework. Still, this article would be the wrong context to attempt such a unification. Methodologically, we determined the scope of the epistemological analysis by a structured literature search, detailed in the Appendix. Our findings partition the epistemology of data science into five areas (see Figure 1), and the article is structured accordingly. Section 2 focuses on descriptive and normative accounts of the composition of data science, i.e., accounts of what data scientists do and should do. Section 3 analyses reflections upon the kind of enquiry that data science is. Section 4 discusses questions about the nature and genealogy of the knowledge that data science produces. Section 5 concentrates on so-called "black box" problems, such as interpretability and explainability. Section 6 explores the epistemically revolutionary new frontier raised by data science: the so-called "theory-free" paradigm in scientific methodology. Section 7 briefly concludes our analysis.

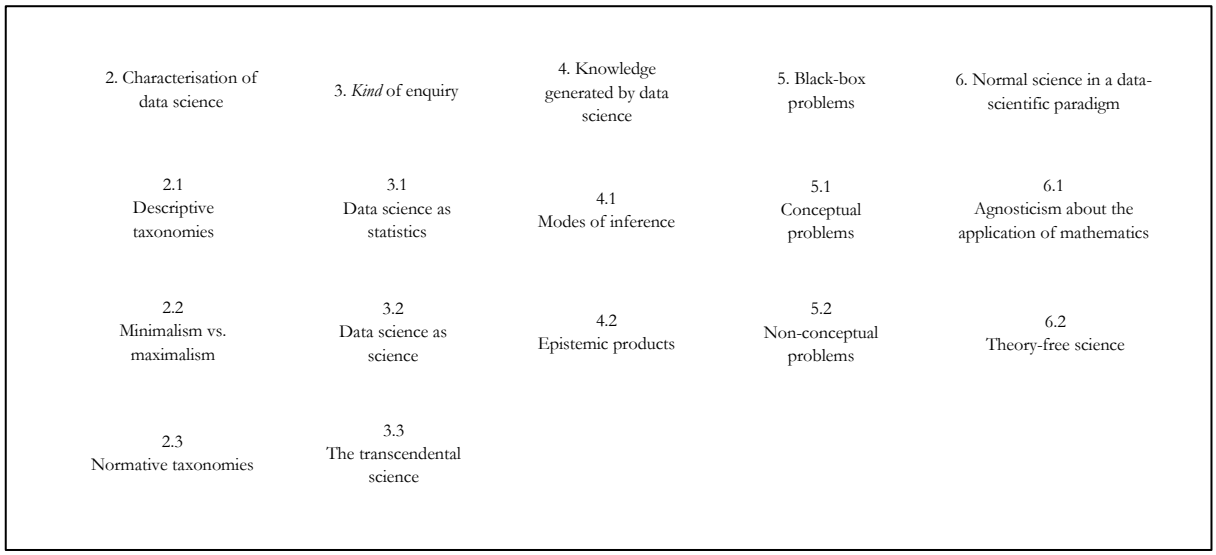

Figure 1. The epistemology of data science: a map.

2. The characterisation of data science

This section reviews the most relevant definitions of data science proposed in the literature, spanning descriptive and normative alternatives. It concludes by offering a new proposal that synthesises the most valuable elements of each.

2.1 Minimalist and maximalist characterisations

The historical origins of data science can be traced back to the work of early naturalists. However, a more recognisable form of the practice began to emerge only with the systematic study of probability and statistics – first through games of chance toward the end of the Renaissance (Hacking, 1975) and later through sociological analyses around the time of the Industrial Revolution (Gigerenzer et al., 2013), culminating in the emergence of genetics in late Victorian Britain (MacKenzie, 1984). Economic incentives were paramount at every turn, be it through better gambling strategies, more accurate actuarial tables, or improved agricultural yields. At the dawn of the twentieth century, statistics came to be recognised as an academic discipline worthy of its own journals and university departments. Technological advances in subsequent decades marked a definite break from theory-driven and inferential classical statistics. New approaches, such as bootstrapping and Markov chain Monte Carlo simulations, replaced strong parametric assumptions with brute computational power. Viewed from this perspective, machine learning algorithms – which automatically detect and exploit subtle patterns in large datasets – are simply the next logical step in a centuries-long progression toward ever more automated forms of empirical reasoning.

The question of when precisely these early forays into quantitative modes of analysis crystallised into what we now call "data science" presupposes that the discipline has some as yet unspecified essential character. Although we are sceptical of any purported "solution" to the so-called "problem" of demarcation – in this area, as in science more generally – we observe two broad trends in the literature on this topic, which we shall deem the "minimalist" and "maximalist" accounts (more on this below). Minimalists aim for necessary conditions, as weakly constraining as possible but still carving out a unique space for data science. Maximalists strive for sufficient conditions with detailed ontologies and methodological taxonomies. Minimalist approaches characterise early debates on the nature of data science. Contemporary analyses tend to embrace maximalist approaches, identifying in data science a means to develop causal knowledge directly connected to the object of analysis.

Minimalist conceptions do not commit data science to any method or subject(s), and do not make any specific claims about what kind of discipline data science is. They focus only on the pedagogical aspects and their dependency on information and data. Chambers (1993) and Carmichael and Marron (2018) provide two examples of minimalist accounts. Chambers (1993) presents a "greater statistics" view of data science, characterised as "everything related to *learning from data*" (Chambers, 1993, p. 182, italics in the original). Similarly, Carmichael and Marron (2018, p. 117) claim that data science is "the business of learning from data" and that a data scientist is someone who "uses data to solve problems".

Maximalist accounts are more fine-grained. Breiman (2001) characterises data science by the kind of knowledge that it generates. Statisticians (taken to include data scientists) may be interested in making correlative predictions from data and extracting information about any associated underlying natural causal mechanisms. Correlative/predictive and causal knowledge are

distinguished and implicitly valued. Causal knowledge is assumed to correspond directly with 'underlying' and 'natural' mechanisms in the real world, while correlative/predictive knowledge may only obliquely correspond to reality, through association with causation. Three important examples in the literature help to clarify this point.

Consider first the maximalist account provided by Mallows (2006, p. 322). It concerns a practical end. As he writes, "Statistics concerns the relation of quantitative data to a real-world problem, often in the presence of variability and uncertainty. It attempts to make precise and explicit what the data has to say about the problem of interest." Mallows emphasises the primacy of problem-solving rather than general pedagogy, a point also stressed by Blei and Smyth (2017) below. A unique aspect of Mallows' account is the explicit mention of variability and uncertainty, which data-scientific and statistical methods must confront. This constitutes an implicit commitment to a separation of the noisy real world and the idealised constructs familiar to the natural and social sciences. Further, statistics is characterised as a fundamental epistemic method in its own right – as *the* bridge between the two worlds.

The second example is offered by Donoho (2017, p. 746), who also supports a maximalist approach. This account of data science has a sociological dimension, referencing the Data Science Association's "professional code of conduct": "'Data Scientist' means a professional who uses scientific methods to *liberate* and *create* meaning from *raw* data [our italics]." This emphasises a close connection between data analysis and scientific inquiry, not just in methodology but also in its fundamental assumptions and aims. The usage of 'liberate' supposes that data originates from processes amenable to systematic study and comprehension. However, the term 'create' implies a deviation from the classical scientific endeavour, suggesting the permissibility of superimposing artificial ontologies upon data as means to whatever ends. Hence data science is carved out as, at least, a quasi-science, if not a full-blown one (more on this in Section 3). It is worth noting that the concept of "raw" data is problematic because data are never entirely devoid of interpretation. As Donoho writes from within the era of big data, his assumption that "raw data" is a suitable base from which to distil and create meaning may be a consequence of the contemporary attitude that data can, are, and will be recorded in sufficient depth, breadth, and quality for any problem domain.

Finally, Blei and Smyth (2017, p. 8691) give a characterisation between minimalism and maximalism: "data science blends statistical and computational thinking… It connects statistical models and computational methods to solve discipline-specific problems." This view commits data science solely to statistical and computational methods, emphasising a practical rather than pedagogical priority. However, this characterisation does not specify information – broadly conceived – as data science's object of interest, nor does it mark specific disciplines as parents or patients of data science.

2.2 Descriptive taxonomies

Some authors have attempted to characterise data science by providing descriptive, procedural taxonomies of the discipline. Three descriptive accounts, written at different times over the last six decades, offer a diachronic perspective.

To our knowledge, Tukey (1962) gave the first descriptive taxonomy of "data analysis" focusing on: "procedures for analysing data and techniques for interpreting the results of such procedures; ways of planning the gathering of data to make its analysis easier, more precise, or

more accurate; all the machinery and results of (mathematical) statistics which apply when analysing data" (Tukey, 1962, p. 2). Tukey intended to give a transparent description of what actually occurs in the analysis of data. As we shall discuss in Section 3, the orthodox view at his time of writing was that data analysis was applied statistics, and hence primarily mathematical. By describing its nature plainly and accurately, Tukey's account was a transgression of the *status quo*: breaking off the concept of data analysis from applied statistics into its own field.

Some years after Tukey, Wu (1997) presented a threefold descriptive taxonomy centred on data collection (experimental design, sample surveys); data modelling and analysis; and problem understanding/solving, and decision making. Like Tukey's, this description came as part of a broader project to move mathematical statistics in a scientific direction. Wu himself bid to rename "statistics" as "data science" or "statistical science", and we note the inclusion of the manifestly scientific "experimental design".

More recently, Donoho (2017) has provided an extensive taxonomy which cites the University of Michigan's "Data Science Initiative" programme: "[Data Science] involves the collection, management, processing, analysis, visualization, and interpretation of vast amounts of heterogeneous data associated with a diverse array of scientific, translational, and interdisciplinary applications" (Donoho, 2017, p. 745). A brief comparison with Tukey's and Wu's accounts highlights the maturation and growth of data science: the procedural pipelines have coevolved with intermediary stages between inputs and products.

Earlier accounts ought not to be faulted for missing a moving target, as they may not have foreseen the growing demands and affordances of the digital era. However, we may still identify a trade-off between constraint specificity and contemporaneity in descriptivist accounts of data science, which has evolved along with computation. Any account that isolates data science from its computational context risks obsolescence, but any account that does not must grapple with a massively entangled and evolving digital sphere, with all its attendant mechanisms and requirements. Therefore, going forward, we distinguish ones-and-zeroes data science from pen-and-paper statistics by its digital and computationally intensive nature.

2.3 Normative Taxonomies

Others thought that the status quo conception of data science of their time was inadequate to meet the demands placed on society by the proliferation of data. This led them to develop revisionist accounts, often proposing normative taxonomies of data science. Here, we consider what may be seen as the four most influential revisionary accounts offered so far: Chambers (1993), Breiman (2001), Cleveland (2001), and Donoho (2017).

Chambers (1993) remarks that, at the time of his analysis, there was a trend in academic statistics towards what he calls "lesser statistics" – mathematical statistics filtered through journals, textbooks, and conferences – rather than towards engaging in real-world applications to data. In this context, he presents the following tripartite taxonomy (Chambers 1993, p. 182) of the composition of his "greater statistics", referring to the concept mentioned earlier as "everything related to learning from data":

1. Preparing data (planning, collection, organization, and validation);
2. Analysing data (by models or other summaries);
3. Presenting data (in written, graphical or other form).

Chambers' taxonomy delineates the processes and products of data science from the decision-making and outcomes that result from those products. The promotion of data preparation to stand equal to analysis and presentation is remarkably prescient. The subprocesses of planning, collection, organisation, and validation anticipate, respectively, the sourcing, volume, diversity, and quality of data required of practical data science, as opposed to the abstract concerns of "lesser statistics". When taken together with the descriptions of the analysis and presentation of data, a conception is revealed of human limitations when confronted with data, and with data science seen as *the* epistemic endeavour to exceed those limitations.

Breiman (2001) echoes the need for statistics to move towards the real world. Like some of the maximalist statements of data science analysed in Section 2.1, his account too emphasises that data analysis collaborates with, and thus acts on, specific disciplines, supplying them with analytical tools. To understand Breiman's radical normative conception of data science as disinterested in truth, in favour of practical knowledge, one needs to engage in a brief historical and sociological detour. In homage to C. P. Snow, Breiman remarks that preference for truth or action characterises two contrasting "cultures" in statistics: the predictive camp, which he estimated at his time of writing in 2001 contained only ~2% of academic statisticians and data analysts, and the inference camp containing the rest. Those in the former camp are primarily interested in generating accurate labels on unseen data. Those in the latter focus on revealing mechanisms and estimating parameters. This is a distinction we shall revisit in Section 4. Breiman's revisionism becomes manifest when he argues that the emphasis on inference over prediction has led to a distracting obsession with "irrelevant theory" and the drawing of "questionable conclusions", thereby keeping statisticians "from working on a large range of interesting current problems". Today, the relative sizes of the two cultures are nearly reversed (e.g., Anderson, 2008). Breiman's vision of a theory-free data science marks a significant deviation from the classic epistemological project of "understanding understanding".

Cleveland (2001, pp. 22-23) considered the teaching programs of his time to be deficient, producing data practitioners unprepared for the demands of an increasingly data-rich society. In this sociological context, he proposes the following taxonomy:

1. Multidisciplinary investigations (data analysis in the context of different discipline-specific areas)
2. Models and methods for data (statistical models, model-building methods, estimation methods, etc.)
3. Computing with data (hardware, software, algorithms)
4. Pedagogy (curriculum planning, school/college/corporate training)
5. Tool evaluation (descriptive and revisionary analysis of tools and their methods of development)
6. Theory (foundational and theoretical problems in data science)

Cleveland's taxonomy puts forward a conception of data science as a fundamentally computational, fully-fledged scientific discipline in its own right. Four points are particularly relevant in this case. First, the elevation of "Computing" alongside "Models and methods for data" marks data science as fundamentally digital, separating it from statistics at large. In contrast to Chambers' taxonomy, computers are by now explicitly recognised as the vehicle that makes data science possible. Second, under "Pedagogy", there is recognition of the necessity to preserve and

propagate data science as an academic and commercial field. Third, the novel inclusion of "Tool evaluation" and "Theory", absent in previous accounts, signals a conception of data science as self-reflective and progressive. Fourth, the fact that "Multidisciplinary investigations" is placed on the same footing as the other five taxa indicates a relative deprioritisation of application. This is a significant shift from preceding accounts, like Breiman's, that treat data science merely as a means to an end.

More recently, Donoho (2017, p. 755) has given a comprehensive revisionary taxonomy to meet current needs. Emulating Chambers' terminology, "greater data science" is set in contrast to some of the descriptive taxonomies described in Section 2.1, which he calls "lesser data science". Greater data science consists of:

1. Data gathering, preparation and exploration
2. Data representation and transformation
3. Computing with data
4. Data modelling
5. Data visualisation and presentation
6. Science about data science

In contrast to Cleveland's taxonomy, Donoho's focuses just on data science *qua* field and means of enquiry. Two aspects of this taxonomy are epistemologically interesting. First, mirroring Cleveland, the repeated presence of the sixth metascientific category: data science should reflect and conduct science on itself in order to improve and develop. Second, Donoho's description is procedurally complete, beginning at data exploration and gathering, and going through all the analytical steps from origins to final products. This ambitious scope contributes to the normative force of Donoho's proposal.

.We have to the end of Section 2. In light of these previous considerations, it seems reasonable to propose the following definition of data science:

> Data science is the study of information systems (natural or artificial), by probabilistic reasoning (e.g., inference and prediction) implemented with computational tools (e.g., databases and algorithms).

This definition is inclusive enough to cover all instances of machine learning – supervised, unsupervised, and reinforcement learning – as well as more generic procedures that typically fall under the umbrella of statistics, such as scatter plotting to inspect trends and bootstrapping to quantify uncertainty. It may or may not exclude some edge cases, depending on one's interpretation of constituent terms. For instance, it covers deterministic systems if one holds that these are a subset of probabilistic systems. It covers hand-calculated regression models if one holds that human cognition is a kind of computation. Yet these are grey areas, even if the former may be an obvious case of computer science and the latter an obvious case of statistics. Data science stretches across both disciplines, emphasising different aspects.

3. Kind of enquiry

Critics may allege that data science is not an academic discipline, but a set of tools bundled together through pragmatic functions. At issue is whether data are the "right kind of thing" to stand as the subject matter of a discipline. If "data" is a concept too insubstantial or the methods of data science are too heterogeneous, then any attempt to carve out a unified data science seems doomed to fail.[1] However, there is a growing demand for data science, not just in the business world, but also in academia, as evidenced by a proliferation of university courses and programs, specialised teaching positions, dedicated conferences, journals, and industry positions. Therefore, for the sake of argument, let us assume that data science is at the very least on its way to becoming an entrenched and mature discipline. The next question is of what kind. The literature offers three main answers: a sort of academic statistics, where statistics is a formal, theoretical part of applied mathematics; statistics, but appropriately expanded to bring it outside of applied mathematics and into a proto-science; and a full-blown science in itself. We now turn to examine each alternative in detail.

3.1 Data science as statistics

The first two of these answers take data science to be some form of statistics. For example, Donoho (2017, p. 746) provides a comprehensive collection of papers, talks, and blogs whose authors argue that data science simply *is* statistics by a different name. This stance further speciates according to whether one takes statistics to be part of, or separate from, applied mathematics. Arguing for the former case, Wu (1997) cites a dictionary definition of statistics: "the mathematics of the collection, organisation and interpretation of numerical data". This narrow view of data analysis does not have many contemporary proponents. Most of the current literature either accepts that data analysis is part of an extended statistics -- which itself is no longer seen as strictly formal mathematics (cf. Chambers' greater statistics) -- or grants data analysis the status of a standalone field, external but related to statistics, which is considered a narrow part of formal, applied mathematics. Breiman's (2001) and Mallows (2006) take the latter stance, by calling for the expansion of statistics to include scientific elements and engage with real-world disciplines. This does not entail that statistics is itself a full-bodied science. Data analysis, on this view, remains statistics, even though it begins to transcend strictly formal mathematical deductive inference and practices.

3.2 Data science as science

Other authors locate data analysis as a scientific discipline. Carmichael and Marron (2018, p. 120) claim that a manifestly scientific data science is a "reaction to the narrow understanding of *lesser statistics*" [our italics]". There are two main strategies to support the claim that data analysis is a science.

The first is to formulate demarcation criteria for whatever it is we already call science (cf. Popper, 1959), and then show that data science satisfies them. Tukey (1962, p.5) made this attempt, setting out three paradigmatic demarcation criteria for science: "intellectual content", "organization into an understandable form", and "reliance upon the test of experience as the ultimate standard of validity". By running up his contemporary data science against these criteria,

[1] This is an open question in the philosophy of information, which we will not address here, as it is deep enough to warrant its own dedicated investigation. We will, however, note that a sustained attempt at some analysis of the concepts of data and information may be found in Floridi (2010).

Tukey concluded that whatever makes other disciplines scientific also applies to data science. (Donoho 2017) focuses on a paradigmatic scientific feature of a subject: the formulation of empirically accountable questions which are solved through scientifically rigorous techniques. Since there is conceptual room for a field of this nature that operates on data and information, he concludes that there is space for a forthcoming genuine science of data analysis.

However, this first strategy clashes with the heterogeneity of science. The demarcation debate lost steam following Laudan's (1983) decisive critique of Popper's falsificationism. In light of this, the second, alternative strategy is to demonstrate relevant similarities between data science and paradigmatic sciences, and that these similarities warrant an extension of the general concept. For example, Wu (1997) cites a series of important similarities between his descriptive taxonomy of statistics and paradigmatic sciences. These similarities include: the "empirical — physical" approach of statistics, in which we use induction to infer knowledge from observations and deduction to infer implications of theories; the primacy of experimental design and data collection; and the use of Bayesian reasoning to evaluate models and evidence. However, there are notable ways in which data science diverges from paradigmatic sciences. Such dissimilarities include the kind of knowledge it generates (see Section 4), the modes of logical inference by which it proceeds (see Section 4), and the status it endows to hypotheses (see Section 6).

A further dissimilarity may be that data science somehow sits alongside normal sciences, providing them with the tools and resources needed to make more profound, discipline-specific discoveries. If these dissimilarities are regarded as sufficiently significant, it becomes plausible that data science might not be a science at all, or perhaps may be a transcendental science. This is the topic of the next section.

3.3 The Transcendental Science

The debate above overlooks the possibility that data science is neither applied statistics nor science but entirely something else. Wiggins has expressed this thought in private communication with Donoho, claiming that "Data science is not a science… It is a form of engineering, and the doers in this field will define it, not the would-be scientists" (Donoho 2017, p. 764). A similar claim could plausibly be made about computer science, which is rooted in mathematics but sufficiently specialized to constitute its own field of inquiry. As it is evident from the descriptive taxonomies above, there are many relevant similarities between computer science and data science. One interesting possibility would be to cleave both data "science" and computer "science" into non-scientific categories of their own.

The point is sharpened if one sees data science as a *basis* for empirical science. Much like how Wittgenstein came to view philosophy as a set of tools and methods for resolving confusion in other areas like mathematics or psychology, data science may be conceived as serving a transcendental function for the sciences, as the condition for the possibility of empirical inquiry as such. There is nothing fundamentally different between, say, Linnaeus' taxonomies and the hierarchical ontologies familiar to database managers. Tycho Brahe's journals are essentially a high-quality dataset. Newton's laws of motion are an algorithm, obtained from and verified against empirical data, for predicting values for some physical variables based on the values of others. We shall not pursue this approach in this article. We posit it more as a suggestion to be explored than a thesis to be defended.

4 The knowledge generated by data science

This section examines the knowledge generated by data science. The analysis is structured into two related parts: the process, or *how,* (concerning modes of inference) and the product, or *what,* (referring to the epistemic products) of data science.

4.1 Modes of inference

Different means of enquiry have differing affinities to the three typical modes of inference: deduction, induction, and abduction. The epistemology of data science reflects on the extent to which data scientists deploy these various modes.

Deductive inferences are present in data science through the heavy reliance on mathematical and logical reasoning. Probability theory, differential calculus, functional analysis, and theoretical computer science are all purely deductive disciplines widely used to derive the properties of algorithms and design new learning procedures with little concern for empirical behaviour. For instance, the backpropagation algorithm used to optimise parameters in neural networks combines elements of linear algebra and multivariable calculus to converge, provably, on a local optimum of an objective function. No datasets are required to derive this result.

Inductive inferences are also of central importance. Data is a finite sample of the world. Data science then identifies structures in the data and distils them into information that applies beyond the data itself. This is achieved by projecting the patterns and structures found in data to new contexts, going beyond the antecedent domain. This projection is precisely inductive inference. This represents a defeasible solution to Hume's problem of induction, whereby statistical testing can provide stronger or weaker evidence in favour of particular hypotheses (Mayo, 1996; 2018). For this reason, Harman & Kulkarni (2007) argue that statistical learning theory represents a principled and sophisticated defence of induction. Similar remarks can be found in Frické (2015), who observes that "Inductive algorithms are a central plank of the Big Data venture." More recently, Schurz (2019) has argued that formal results from reinforcement learning demonstrate the optimality of meta-induction, thereby solving Hume's problem on *a priori* grounds.

One can distinguish between two canonical types of inductive inference, *object* and *rule* induction. The first is the informed prediction of singular unobserved instances: hypotheses of the form "the next observed instance of X will be Y" based on previous data of the co-instantiation of X's and Y's. This is known as *object* induction. *Rule* induction, by contrast, posits universal claims of the form "all X's are Y's", based on the same data. Data-scientific investigation involves both. Singular predictive instances are commonplace in any application of supervised learning, where the goal is to learn a function from inputs to outputs. These are the kinds of inductions that interest Breiman's (2001) "first culture" of statistics. At the same time, one of the purposes of data science is to identify underlying structures and mechanisms. The project of causal inference, which we revisit in Section 4.2, is devoted to such forms of rule induction.

Turning to abductive inference, Alemany Oliver and Vayre (2015) have emphasised the importance of abductive reasoning in data science methods, particularly in how data science is embedded into broader scientific practice (see Section 6 for further discussion). They argue that the tools of data science have an interest first in the exploration of data to determine its internal structure, and second in the identification of the best hypotheses to explain this structure. This inference from structure to an explanatory hypothesis is an abductive inference. The view that

science is essentially abductive can be traced back to Peirce, though modern adherents abound (Harman, 1965; Lipton, 1991; Niiniluoto, 2018). The status of abduction in a data-intensive context is further elevated by the theoretical virtue of explanatory unification (Kitcher, 1989). In the philosophy of science, a common virtue of a theory is its explanatory power, with some authors maintaining that such power is grounds to choose one of two empirically equivalent theories (cf. van Fraassen's (1980) discussion of pragmatic virtues). One dimension of explanatory power is the extent of the diversity and heterogeneity of phenomena that a theory can explain simultaneously (cf. Kitcher, 1976). If the methods of data science allow for the identification of patterns in a diverse and heterogeneous range of phenomena, then perhaps we will develop a broader and more nuanced picture of the explanatory power of our theories. For those theories that can unify many phenomena, abductive reasoning confers more robust support on them considering various data science techniques.

In addition to being an end in itself, epistemological reflection on modes of influence also sheds light on the connections between data science, mathematics, and science. The similarities between these disciplines – such as their relevance, explanatory power, practical utility, and degree of success – are precisely what is in question when we look to extend the categories coherently. For example, mathematical proofs are formulated deductively. But given the importance of non-deductive inferences in data science, one needs to recognise an important dissimilarity between the two and refrain from placing data science strictly within applied mathematics. Likewise, natural sciences use a mixture of deduction, induction, and abduction in their everyday practice, with more formal sciences making more frequent use of deduction, and more applied sciences relying more on abduction. Other sciences assign different weightings to differing modes of inference. For example, abduction is commonplace in the social, political, and economic sciences. Cognitive science is another example that relies on abduction given the frequency of empirically equivalent, underdetermined theories. It seems that data science, if it is a science, is in good company.

4.2 Epistemic products

The trichotomy of machine learning – which spans supervised, unsupervised, and reinforcement learning algorithms – helps to delineate the kind of knowledge generated by data science and its techniques.

Supervised learning models predict outcomes based on observed associations. They automate the process of inductive reasoning at scales and resolutions that far exceed the capacity of humans. However, large datasets and powerful algorithms are not sufficient to overcome the fundamental challenges inherent to this mode of inference. A model that does well in one environment may fail badly in another if data no longer conform to the observed patterns. For instance, a classifier trained to distinguish cows from camels may struggle when presented with a cow in the desert or a camel on grass, presuming the training set only contains images of both animals in their natural habitats. Since the background was a reliable indicator of the outcome in training, the model could be forgiven for assuming the same would hold at test time. This fallibility is inherent in all inductive reasoning, which nevertheless helps us accomplish many important epistemic goals.

Unsupervised learning is a more heterogeneous set of methods, broadly united by their tendency to infer structure without any predefined outcome variable(s). Examples include clustering algorithms, autoencoders, and generative models. At their best, these tools can shed

light on latent properties – how samples or features reflect some underlying facts about the data generating process. For instance, clustering methods are commonly used in cancer research to categorise patients into subgroups based on biomarkers. The idea is that an essential fact (e.g., that cancer manifests in identifiable subtypes) is reflected by some contingent property (e.g., gene expression levels). The risk of overfitting – i.e., "discovering" some structure in training data that does not generalise to test data – is especially high in this setting, as there is no outcome variable against which to evaluate results.

In reinforcement learning, one or more agent(s) must navigate an environment with little guidance beyond a potentially intermittent reward signal. The goal is to infer a policy that maximises rewards and/or minimises costs. A good example of this is the multi-armed bandit problem. An agent must choose among a predefined set of possible actions – i.e., must "pull" some "arm" – without knowing the rewards or penalties associated with each. Therefore, an agent in this setting has to strike a difficult balance between exploration (randomly pulling new arms) and exploitation (continually pulling the arm with the highest reward thus far). Reinforcement learning has powered some of the most notable achievements of data science in recent years, such as AlphaGo, an algorithm that is currently the world's best player of Go, chess, and several other classic board games. The epistemic product of such algorithms is neither associations (as in supervised learning) nor structures (as in unsupervised learning), but *policies* – methods for making decisions under uncertainty.

On their own, these methods do not necessarily provide causal knowledge. However, some of the most important research on AI of the last 20 years has focused on causal reasoning (Spirtes et al., 2000; Pearl, 2009; Imbens & Rubin, 2015; Peters et al., 2017). Such research demonstrates how probabilistic assumptions can combine with observational and/or interventional data to infer causal structure and treatment effects. Remarkably, this literature is only just beginning to gain traction in the machine learning community. Recent work in supervised learning has shown how causal principles can improve out-of-distribution performance (Arjovsky et al., 2019), while complex algorithms such as neural networks and gradient boosted forests are increasingly used to infer treatment effects in a wide range of settings (Nie and Wager, 2021). Causal discovery – the task of learning causal associations from observational data – is a quintessential unsupervised learning problem. This has been an active area of research since at least the 1990s and remains so today (see Glymour et al. (2019) for a recent review). Reinforcement learning – perhaps the most obviously causal of all three branches, given its reliance on interventions – has been the subject of intense research in the last few years (Bareinboim et al., 2021). Various authors have shown how causal information can improve the performance of these algorithms, which in turn helps reveal causal structure.

These methods can, in principle, be used to infer natural laws. Schmidt and Lipson (2009) have proposed what appears to be the algorithmically obtained laws of classical mechanics. Their method involved analysing the motion-data of various dynamical systems using algorithms that had no prior physical knowledge of mechanics. They claim to obtain the Lagrangian and Hamiltonian of those dynamical systems, together with various conservation laws. This provides an attractive data point for those who are hopeful for the possibility of the autonomous discovery of natural laws. The roles of correlation and causation in science, and of autonomous, theory-free science are discussed in Section 6.

5. Black box problems

The tools of data science have become highly sophisticated and complex. This is partly because data science has always been accountable to practical motivations. Any development that produces a more successful (more efficient, accurate, deployable, etc.) outcome is adopted in virtue of its utility, often without pausing for reflection on how it is to be embedded in our wider conceptual schemes. This has led to questions about the opacity of these tools. In this section, we evaluate various types of black box problems proposed in the literature.

First, it may be helpful to provide a clarification. Burrell (2016) has proposed that there are three ways in which data science algorithms become opaque. The first is their intentional concealment for commercial or personal gain. The second is the opacity that arises from technological literacy and proficiency being a necessary condition to understand sophisticated algorithms. And the third is inherent complexity arising from algorithmic optimisation procedures that exceed the capacity of human cognition. The first two of these problems are pragmatic problems that occur when data science is embedded in wider society (see Tsamados et. al. (2021) for recent work on these issues). They are not the kind of epistemological problems with which we are concerned here. Thus, we will focus only on the last problem. Furthermore, there have been many technical solutions or proto-solutions to various instances of black box problems. The nature of these solutions does not concern us because we are focused on the philosophical level of abstraction above such technical investigations. In this section, we provide only a brief, comparative overview in order to illustrate (dis)similarities, or instances where putatively different problems may collapse into one.

To begin with, black box problems fall into one of two kinds, which we call *conceptual* and *non-conceptual*. Conceptual problems are those which concern the boundaries of the concepts that are employed in discussing black boxes. For example, whether simply trying to project concepts like "explainability" into a machine learning context can be achieved in a coherent and non-ambiguous way. Non-conceptual problems, conversely, do not concern the nature, boundaries, and coherences of employed concepts themselves, but the broader problems that result from the use of these concepts, such as in epistemology. Here, we will restrict our focus only to non-conceptual problems in the domain of epistemology. However, it is worth bearing in mind that further non-conceptual problems arise elsewhere, for example, in ethics or politics.

5.1. Conceptual problems

Some black box problems arise from our ordinary concepts being in some way inadequate or unclear when projected into machine-learning contexts. Lipton (2018) has acknowledged this imprecision over the use of "interpretation". He observes that "the task of interpretation appears underspecified. Papers provide diverse and sometimes non-overlapping motivations for interpretability and offer myriad notions of what attributes render models interpretable" (Lipton, 2018, p. 36). Similarly, Doshi-Velez and Kim (2017) have remarked on the lack of agreement of a definition of "interpretability", and further about how it is to be evaluated. They identify two paradigmatic uses of "interpretability" in the literature: interpretability in the context of an application and interpretability through a quantitative proxy. Rigorous definitions of both are found lacking.

There have been a few attempts to respond to such conceptual problems. One important first step is to construct a clear taxonomy of how problematic concepts like interpretability are

used, and what the desiderata and methodologies of interpretability research are. This is the kind of project in which Lipton (2018) engages. A second attempt is to refine these concepts, or at least conduct the groundwork to facilitate this refinement. Doshi-Velez and Kim (2017) engage in this kind of project, laying the groundwork for the subsequent rigorous definition and evaluation of interpretability. Authors have also refined the concepts of interpretability and its cognates by making fine-grained distinctions within them, adding to their structure. Doshi-Velez and Kim (2017) distinguish between local and global interpretability to avoid confusion. The former applies to individual predictions and the latter to the entire decision boundary or regression surface. Watson and Floridi (2020) make a similar distinction between local (token) and global (type) explanations, though in a more formal mathematical context.

Further work on the representations deployed in black box problems concerns the relationship between various roughly synonymous terms: words like "interpretability", "explainability", "understandability", "opacity", and so on. It is of philosophical interest whether any or all of these terms overlap in whole or in part. Some commentators take a coarse-grained approach to such cognates. Krishnan (2020), for example, takes them as negligibly different, arguing that these terms all define one another in a circular fashion that does little to clarify imprecise concepts. Others take a more fine-grained approach. Tsamados et al. (2021) emphasise a difference between explainability and interpretability. The former applies to experts and non-experts alike, for example, the expert data scientist practitioner might need to explain the mechanics of some algorithm to their non-expert client. In contrast, the latter is restricted to experts (interpretability as interpretability-in-principle). Thus, in their view, explainability presupposes interpretability but not vice versa.

5.2 Non-conceptual problems

Non-conceptual problems and their solutions do not address deficiencies in representations themselves. In this section, we will discuss four epistemological problems that have received less attention.

Ratti and López-Rubio (2018) have argued that interpretability is crucial to distil causal explanations from the correlations identified by data science techniques, as may be the case in a data-rich scientific context. Through the paradigm of mechanistic biological models, they observe that for biologists to turn data-scientific correlative models into causal models with explanatory power, the correlative models must be interpretable. This stems from a general trade-off: the more complex a model is, the less explanatory it is. Since the predictive powers of data-scientific models are positively correlated with their complexity, they conclude that there is a genuine epistemological black box problem.

Watson and Floridi (2020) have construed overfitting as a different kind of epistemological black box problem, as a kind of algorithmic Gettier case. Overfitting occurs when a machine learning model makes correct predictions in the training corpus but fails to predict correctly in testing data. They cite the results of Lapushkin et al. (2016), in which pictures of a horse shared a subtle, distinctive watermark. The resultant image classifier strongly associated that watermark with the label "horse", and thus could not correctly classify horses in a test set when the watermark was absent. This, Watson and Floridi propose, is similar to Gettier cases, where one forms an accidentally true belief through unreliable knowledge-generating mechanisms. They argue that a framework for interpreting black boxes will reduce overfitting.

Krishnan (2020) has remarked on the broader epistemological point that, insofar as machine learning algorithms might have a pedagogical dimension (that we can learn from the mistakes that algorithms might make), they must be interpretable or understandable for us to learn anything at all. Lipton (2018, section 2.4) (citing Kim et al., 2015 and Huysmans et al., 2011) makes a similar remark on the informativeness of algorithms. Thus, there are significant epistemic benefits to greater algorithmic transparency.

The discussion above gives the impression that these problems are substantial and worth solving. However, not all commentators agree. There are two main kinds of objections. Some concede that black boxes are opaque but deny that the correct way to proceed is to try to explain or interpret their inner workings. Instead, they argue that black boxes should be replaced altogether by equally capable non-black boxes. Others deny that black boxes are problematic at all. We will present both sorts in turn.

Rudin (2019) has expressed an objection of the first kind. She agrees that the lack of interpretability of machine learning algorithms is a problem. However, she takes this not as motivation to construct better *post hoc* interpretability methods, but instead as a reason to reject opaque models altogether. She rejects the commonplace assumption that accuracy and interpretability are inversely related. In her view, black box problems are to be dissolved (rather than solved) by globally transparent models that perform comparably to black box competitors.

Zerelli et al. (2019) have expressed an objection of the second kind, arguing that the opacity of black boxes is not a genuine problem at all. They see the explainability debate as evidence of a pernicious double standard. They point out that we do not demand explicit, transparent explanations from human judges, doctors, military generals, or bankers. Rather, justification is found simply in past reliability: demonstrated and sustained accuracy and success. If we impose the same norms on algorithms, then the explainability problem is once again dissolved.

Along similar lines, Krishnan (2020) has argued that our concerns about interpretability and its cognates are unnecessarily inflated. The inherent imprecision of these terms prevents them from doing the work required of them: "Interpretability and its cognates are unclear notions… We do not yet have a grasp of what concept(s) any technical definitions are supposed to capture — or indeed, whether there is any concept of interpretation or interpretability to be technically captured at all" (Krishnan, 2020, p. 490). But unlike Doshi-Velez and Kim, Krishnan does not take this as motivation to sharpen such concepts for subsequent progress, for worrying about them distracts from our real needs. Krishnan contends that most of the *de facto* motivations for treating interpretability as an epistemological problem in the first place are due to other ends (e.g., social, political, etc.). For example, algorithmic bias audits use explainability as a means to avoid unethical consequences.

We are sympathetic to Krishnan's overall project. Many authors uncritically assume that black box problems are necessarily important, and epistemological concerns about concepts like interpretability are often in practice means to other ends. However, we disagree that these exhaust the epistemological utility of such concepts, as the examples from Section 5.2 attest. It might be the case that worrying about black box problems is an inefficient and suboptimal use of philosophical effort (particularly in the hyper-pragmatic context in which data science methods are mostly deployed). However, black box problems *qua* objects of epistemological interest remain relevant to at least some parts of a complete philosophy of data science.

6. Normal science in a data-intensive paradigm

Having so far considered foundational issues in the philosophy of data science, we may now broaden the investigation to consider how data science might shape science and the philosophy of science in general. Kuhn (1970) famously proposed that science goes through cycles of normality, crisis, and ultimately revolution. The normal phase features practitioners engaged in the gnostic pursuit of puzzle-solving using the tools of the prevailing paradigm. However, it has recently been proposed that the proliferation of data has inaugurated a new era of agnostic science. Here, scientific knowledge can be generated, and mathematical and data-scientific methods deployed, without any prior knowledge or understanding of phenomena or their interrelations. Kitchin (2014) has compiled Gray's work (found in Hey et al. (2009)) to elucidate the nature of this new paradigm and locate it in the history of science (see Table 1). This section explores the extent and the implications of agnostic science.

Paradigm	Nature	Form	When
First	Experimental	Empiricism; describing natural phenomena	pre-Renaissance
Second	Theoretical	Modelling and generalisation	pre-computers
Third	Computational	Simulation of complex phenomena	pre-Big Data
Fourth	Exploratory	Data-intensive; statistical exploration and data mining	Now

Table 1. Scientific paradigms (taken from Kitchen (2014, p. 3), compiled from Hey et al. 2009)

6.1. Agnosticism about the application of mathematics

One identification of agnosticism is provided by Napoletani et al. (2018), who observe that the *de facto* application of mathematical techniques in science is undergoing an agnostic transformation. They remark that classical methods required both the prior understanding of phenomena and interconnections between elements in datasets. This is the case, for example, if one wishes to model some biological population using differential equations. The nature of the models one uses, which parameters to include, and so on, require the scientist to have antecedent knowledge and understanding about population biology, multivariate calculus, etc. They also need the scientist to know the basic structure of the dataset. Matters are very different in contemporary data analysis. There, the scientist can remain to a great extent agnostic or uninformed about any underlying scientific theory and the structure of their data. With the tools of contemporary data science, raw data can be parsed, and structure exploited more or less automatically.

After observing that this appears to be an important direction in scientific practice, Napoletani et al. raise the second-order question of why mathematics and data have such an effective synergy. They claim that a common response is to appeal to a Wignerian-like resignation to "unreasonable effectiveness" (Wigner, 1960). On this view, big data has a sort of omnipotence that grants unreasonable success to disparate and heterogenous data-scientific tools. However, Napoletani et al. reject this response, arguing that the question can be reformulated into the more general question of whether the success of mathematical methods in an agnostic normal science is due to a similarity between the structure of those methods and the structure of the phenomena

themselves captured in data corpora. This is a question that deserves further attention in the debate.

6.2. Theory-free science

While Napoletani et al. observe the increasing possibility of employing mathematical techniques agnostically, others have engaged in a more radical debate about whether this agnosticism heralds the end of theory choice in science altogether. Anderson (2008) has argued that classical theory-driven science is becoming obsolete. In his view, the density and plurality of correlations yielded by the analysis of extraordinary large amounts of data will become more useful than the causal generalisations provided by classical science. Other authors have made similar remarks (Prensky, 2009; Steadman, 2013). Kitchin (2014) provides a more formal characterisation of this view, which he calls a new type of empiricism. Schmidt and Lipson's (2009) aforementioned reconstruction of classical mechanics via machine learning is a provocative example of theory-free science in action.

Critics object that this is sensationalist, over-optimistic and inflated. Kitchin (2014) presents a fourfold attack on Schmidt and Lipson's (2009) analysis. His first contention is that, as much as large data corpora can try to exhaust information in a whole domain, they are nonetheless coloured by the technology used in their generation and manipulation, the data ontology in which they exist, and the possibility of sampling bias. Indeed, "all data provide oligoptic views of the world" (Kitchin, 2014). Second, following Leonelli (2012), he remarks that even the agnostic distillation of structure and patterns from data cannot occur in vacuo from all scientific theory. Due to their deep embedding in society, scientific theories and training always provide the scaffolding around data collection and analysis. Third, insofar as normal science is cumulative, he argues that the individual results of data-scientific investigations will always require interpretation and framing by scientists who themselves are equipped with knowledge of scientific theories. And fourth, if data and the results of its analysis are interpreted free of any background theory, they risk becoming fruitless. It will be difficult for them to contribute to any fundamental understanding of the nature of phenomena since it "lacks embedding in wider … knowledge" (Kitchin, 2014). Frické (2015) presents a similar view against this extreme kind of agnosticism. He objects that one needs antecedent theoretical insight to decide which data to provide inductive algorithms in the first place. Theory cannot be removed from science, even in a data-driven paradigm.

We believe that these arguments can be supplemented with two further reasons against total agnosticism. The first relates to the critical issue in the philosophy of science of the theory-ladenness of observation, which holds that what one observes is influenced by one's theoretical and pre-theoretical commitments. This is doubly true for data science, where observations are gathered, labelled, and processed according to pre-existing categories and analysis routines. Second, it is plausible that Anderson's claim that correlations will be sufficient for the future of science is too naïve a conception of the scientific enterprise. It reminds one of Francis Bacon's untenable view that Nature would speak by itself if adequately interrogated. Agnostic data science may indeed generate a predictive science without knowledge of any underlying natural laws or causal mechanisms. But prediction is not the only goal of the scientific enterprise — another is to explain phenomena through coming to know the underlying causal structure of the world, which helps to plan and intervene.

Total agnosticism, therefore, seems too extreme. The task then is how to integrate agnostic data-scientific practices into scientific methodology. Kitchin (2014) proposes a humbler account of this integration. He calls it "data driven science", and it takes the form of a rebalancing of the

three modes of inference discussed in Section 4.1. He argues that contemporary normal science has an experimental-deductive dimension in which hypotheses are deduced from more fundamental hypotheses and then offered up for confirmation or refutation by experiment. In contrast, science in a data-driven paradigm elevates the status of inductive logic in this process of hypothesis formation, with experimental hypotheses generated from correlations identified by data-scientific methods rather than by deduction from parent hypotheses. However, in contrast to the naïve empiricist, Kitchin's data-driven science does not involve the absolute primacy of induction. Theories and their deductions play an essential role, for example, in framing data, directing which data-scientific processes to deploy, embedding results in wider knowledge, generating causal explanations, and so on. A picture of a new science emerges, involving a shift towards a more inductive enterprise, while maintaining many paradigmatic and realistic similarities to our current model of normal science.

There have been further remarks about the introduction of data-scientific methods into the social sciences. Lazer et al. (2009) stress the emergence of "computational social science", and Miller (2010) observes the proliferation of data in the context of regional and urban science. In both cases, the potential for data to reshape social-scientific practices is acknowledged. However, authors have noted the dissimilarities between natural and social sciences, which likely mean that the impact of data on the two categories will differ.

It is likely that the future of data-intensive science will still be theory-based, though sometimes agnostic and data-scientific methods to assist in theory-generation will be used. Since Reichenbach (1938), there has been a popular distinction made in the philosophy of science between the context of *discovery* and the context of *justification*: where a theory came from is irrelevant to whether the theory is sound. Consequently, it has become orthodox to consider scientific theories only for their own content, independent of their origins. The genealogy of our scientific knowledge has, classically, never been of epistemic relevance.

This distinction may be brought into question with the possibility of agnostic science in a data-intensive paradigm. For now, the genealogy of such agnostic knowledge that is generated autonomously from data is important: its epistemic standing is supervenient on the tools and algorithms of data science that generated it and on the quality of the antecedent data. Thus, the reliability of automated inferences depends on the quality of the underlying data and the algorithm(s) used to extract information from them. Such questions about theory genealogy are perhaps too often ignored by modern philosophies of science that inform "gnostic" paradigms. A philosophy of science in a data-intensive paradigm may be forced to address them more directly.

7. Conclusion

In this article, we provide a systematic and integrated analysis of the current landscape of the epistemology of data science. We have focused on its critical evaluation and identifying and characterising some of its pressing or obvious gaps wherein philosophical interest lies. We have structured this reconstruction into five areas: descriptive and normative accounts of the composition of data science; reflections upon the kind of enquiry that data science is; the nature and genealogy of the knowledge that data science produces; "black box" problems; and the nature and standing of a new frontier within the philosophy of science that is raised by data science. Each of these areas is home to a variety of important issues and active debates, and each area interacts with the others. The resulting picture is a rich, interconnected, and flourishing epistemology, which

will doubtlessly continue to expand as both philosophical and technological progress is made, and possibly influence other interconnected views about the nature of science and its foundations.

Appendix: Details of literature search

The literature search for the present work was conducted as described by the following Table 2:

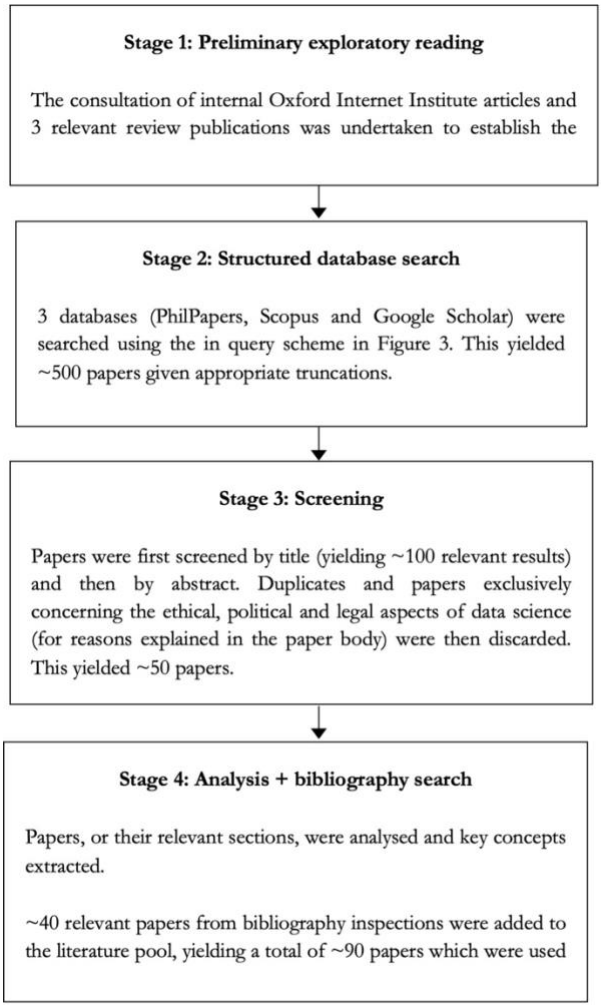

Table 2. Scheme of literature search

Literature was restricted exclusively to papers written in English. The impact of this choice on the analysis is likely minimal as we do not expect a considerable geographic, cultural, or linguistic variation in foundational questions in data science, given the global, highly interdisciplinary, and contemporary nature of the discipline. This is, perhaps, unlike other philosophical issues (e.g., those ethical), which may be more sensitive to such variation in genealogy or circumstance.

Database	Search Query
Philpapers	"data science"
	"data science" & "epistemology"
	"big data"
Scopus, Google Scholar	"data science" & "epistemology"
	"data science" & "philosophy"
	"big data" & "epistemology"
	"big data" & "philosophy"

Figure 3. Table of search queries

Bibliography

Alemany Oliver, M. and Vayre, J.-S. (2015) 'Big data and the future of knowledge production in marketing research: Ethics, digital traces, and abductive reasoning', *Journal of Marketing Analytics*, 3(1), pp. 5–13. doi: 10.1057/jma.2015.1.

Anderson, C. (2008) 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete', *Wired*. Available at: https://www.wired.com/2008/06/pb-theory/ (Accessed: 14 December 2020).

Arjovsky, M., Bottou, L., Gulrajani, I., & Lopez-Pad, D. (2019) Invariant risk minimization. *arXiv* preprint, 1907.02893.

Baker, M. (2016) '1,500 scientists lift the lid on reproducibility', *Nature News*, 533(7604), p. 452. doi: 10.1038/533452a.

Bareinboim, E. and Pearl, J. (2016) 'Causal inference and the data-fusion problem', *Proceedings of the National Academy of Sciences of the United States of America*, 113(27), pp. 7345–7352. doi: 10.1073/pnas.1510507113.

Bareinboim, E., Lee, S., & Zhang, J. (2021) An introduction to causal reinforcement learning. Columbia CausalAI Laboratory, Technical Report (R-65).

Blei, D. M. and Smyth, P. (2017) 'Science and data science', *Proceedings of the National Academy of Sciences*, 114(33), pp. 8689–8692. doi: 10.1073/pnas.1702076114.

Breiman, L. (2001) 'Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author)', *Statistical Science*, 16(3), pp. 199–231. doi: 10.1214/ss/1009213726.

Burrell, J. (2016) 'How the machine "thinks": Understanding opacity in machine learning algorithms:', *Big Data & Society*. doi: 10.1177/2053951715622512.

Canali, S. (2016) 'Big Data, epistemology and causality: Knowledge in and knowledge out in EXPOsOMICS', *Big Data and Society*, 3(2).

Carabantes, M. (2020) 'Black-Box Artificial Intelligence: An Epistemological and Critical Analysis', *AI and Society*, 35(2), pp. 309–317. doi: 10.1007/s00146-019-00888-w.

Carmichael, I. and Marron, J. S. (2018) 'Data Science vs. Statistics: Two Cultures?', *Japanese Journal of Statistics and Data Science*, 1(1), pp. 117–138. doi: 10.1007/s42081-018-0009-3.

Carroll, J. W. (2020) 'Laws of Nature', in Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy*. Winter 2020. Metaphysics Research Lab, Stanford University. Available at: https://plato.stanford.edu/archives/win2020/entries/laws-of-nature/ (Accessed: 18 December 2020).

Chambers, J. M. (1993) 'Greater or lesser statistics: a choice for future research', *Statistics and Computing*, 3(4), pp. 182–184. doi: 10.1007/BF00141776.

Cleveland, W. S. (2001) 'Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics', *International Statistical Review / Revue Internationale de Statistique*, 69(1), pp. 21–26. doi: 10.2307/1403527.

Crawford, K. (2014) 'Critiquing Big Data: Politics, Ethics, Epistemology | Special Section Introduction', p. 10.

Donoho, D. (2017) '50 Years of Data Science'. doi: 10.1080/10618600.2017.1384734.

Doshi-Velez, F. and Kim, B. (2017) 'Towards A Rigorous Science of Interpretable Machine Learning', *arXiv:1702.08608 [cs, stat]*. Available at: http://arxiv.org/abs/1702.08608 (Accessed: 7 December 2020).

Elragal, A. and Klischewski, R. (2017) 'Theory-driven or process-driven prediction? Epistemological challenges of big data analytics', *Journal of Big Data*, 4(1), p. 19. doi: 10.1186/s40537-017-0079-2.

Floridi, L. (2010) Information: A Very Short Introduction. Oxford University Press.

Floridi, L. (2012) 'Big Data and Their Epistemological Challenge', *Philosophy & Technology*, 25(4), pp. 435–437. doi: 10.1007/s13347-012-0093-4.

van Fraassen, B. C. (1980) *The Scientific Image*. Oxford University Press.

van Fraassen, B. C. (1989) *Laws and Symmetry*, *Laws and Symmetry*. Oxford University Press. Available at: https://oxford.universitypressscholarship.com/view/10.1093/0198248601.001.0001/acprof-9780198248606 (Accessed: 14 December 2020).

Frické, M. (2015) 'Big data and its epistemology', Journal of the Association for Information Science and Technology, 66(4), pp. 651–661. doi: 10.1002/asi.23212.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J. & Krüger, L. (2013) *The empire of chance: How probability changed science and everyday life*. New York: Cambridge University Press.

Glymour, C., Zhang, K., & Spirtes, P. (2019) Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, *10*, 524.

Hacking, I. (1975) *The emergence of probability: A philosophical study of early ideas about probability, induction, and statistical inference.* New York: Cambridge University Press.

Hey, T., Tansley, S. and Tolle, K. (2009) 'The Fourth Paradigm: Data-Intensive Scientific Discovery', p. 287.

Harman, G. (1965) The inference to the best explanation. *Philosophical Review*, *74*(1), 88-95.

Harman, G. & Kulkarni, S. (2007) *Reliable reasoning: Induction and statistical learning theory*. Cambridge, MA: The MIT Press.

Hillar, C. and Sommer, F. (2012) Comment on the article 'Distilling free-form natural laws from experimental data'.

Hooker, G. and Hooker, C. (2017) 'Machine Learning and the Future of Realism', *arXiv:1704.04688 [cs, stat]*. Available at: http://arxiv.org/abs/1704.04688 (Accessed: 24 September 2020).

Imbens, G.W. and Rubin, D.B. (2015) *Causal inference for statistics, social, and biomedical sciences: An introduction*. New York, NY, US: Cambridge University Press. doi:10.1017/CBO9781139025751.

Kelling, S. *et al.* (2009) 'Data-intensive Science: A New Paradigm for Biodiversity Studies', *BioScience*, 59(7), pp. 613–620. doi: 10.1525/bio.2009.59.7.12.

Kerridge, I., Mason, P. and Lipworth, W. (2017) 'Ethics and Epistemology of Big Data', *Journal of Bioethical Inquiry*, 14(4), pp. 485–488.

Kim, B. *et al.* (no date) 'iBCM: Interactive Bayesian Case Model Empowering Humans via Intuitive Interaction', p. 12.

Kitchin, R. (2014) 'Big Data, new epistemologies and paradigm shifts', *Big Data & Society*, 1(1), p. 2053951714528481. doi: 10.1177/2053951714528481.

Kitcher, P. (1976) 'Explanation, Conjunction, and Unification', *The Journal of Philosophy*, 73(8), pp. 207–212. doi:10.2307/2025559

Kitcher, P. (1989) Explanatory unification and the causal structure of the world. In P. Kitcher & W. Salmon (eds.), *Scientific Explanation*, pp. 410-505. Minneapolis: University of Minnesota Press.

Krishnan, M. (2020) 'Against Interpretability: a Critical Examination of the Interpretability Problem in Machine Learning', *Philosophy & Technology*, 33(3), pp. 487–502. doi: 10.1007/s13347-019-00372-9.

Kuhn, T. S. (1970) *The structure of scientific revolutions*. 2nd Edition. Chicago: University of Chicago Press.

Laudan, L. (1983) The demise of the demarcation problem. In R.S. Cohen & L. Laudan (Eds.), *Physics, philosophy and psychoanalysis: Essays in honor of Adolf Grünbaum* (pp. 111-127). Dordrecht: Springer.

Lapuschkin, S. *et al.* (2016) 'Analyzing Classifiers: Fisher Vectors and Deep Neural Networks', in. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2912–2920. Available at: https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Bach_Analyzing_Classifiers_Fisher_CVPR_2016_paper.html (Accessed: 17 December 2020).

Lazer, D. *et al.* (2014) 'The Parable of Google Flu: Traps in Big Data Analysis', *Science*, 343(6176), pp. 1203–1205. doi: 10.1126/science.1248506.

Leonelli, S. (2014) 'What difference does quantity make? On the epistemology of Big Data in biology', *Big Data & Society*, 1(1), p. 2053951714534395. doi: 10.1177/2053951714534395.

Lipton, P. (1991). *Inference to the best explanation*. London: Routledge.

Lipton, Z. C. (2018) 'The mythos of model interpretability', *Communications of the ACM*, 61(10), pp. 36–43. doi: 10.1145/3233231.

Lowrie, I. (2017) 'Algorithmic rationality: Epistemology and efficiency in the data sciences', *Big Data & Society*, 4(1), p. 2053951717700925. doi: 10.1177/2053951717700925.

MacKenzie, D. (1984) *Statistics in Britain, 1865-1930: The social construction of scientific knowledge*. Edinburgh: Edinburgh University Press.

Mallows, C. (2006) 'Tukey's Paper After 40 Years', *Technometrics*, 48, pp. 319–325. doi: 10.1198/004017006000000219.

Maruyama, Y. (2021) 'Post-truth AI and big data epistemology: From the genealogy of artificial intelligence to the nature of data science as a new kind of science', *Advances in Intelligent Systems and Computing*, 1181 AISC, pp. 540–549. doi: 10.1007/978-3-030-49342-4_52.

Mayo, D. (1996) *Error and the growth of experimental knowledge.* Chicago: University of Chicago Press.

Mayo, D. (2018) *Statistical inference as severe testing: How to get beyond the statistics wars.* New York: Cambridge University Press.

Mazzocchi, F. (2015) 'Could Big Data be the end of theory in science?', *EMBO reports*, 16(10), pp. 1250–1255. doi: 10.15252/embr.201541001.

Miller, H. J. (2010) 'The Data Avalanche Is Here. Shouldn't We Be Digging?', *Journal of Regional Science*, 50(1), pp. 181–201. doi: https://doi.org/10.1111/j.1467-9787.2009.00641.x.

Mittelstadt, B. D. *et al.* (2016) 'The ethics of algorithms: Mapping the debate', *Big Data & Society*, 3(2), p. 2053951716679679. doi: 10.1177/2053951716679679.

Napoletani, D., Panza, M. and Struppa, D. (2018) 'The Agnostic Structure of Data Science Methods', p. 17.

Neresini, F. (2017) 'On Data, Big Data and Social Research. Is It a Real Revolution?', in Lauro, N. C. et al. (eds) *Data Science and Social Research*. Cham: Springer International Publishing (Studies in Classification, Data Analysis, and Knowledge Organization), pp. 9–16. doi: 10.1007/978-3-319-55477-8_2.

Nie, X. and Wager, S. (2021) 'Quasi-oracle estimation of heterogeneous treatment effects', *Biometrika*, 108(2), pp. 299–319. doi:10.1093/biomet/asaa076.

Niiniluoto, I. (2018) *Truth-seeking by abduction*. Cham, Switzerland: Springer.

Pearl, J. (2009) *Causality*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511803161.

Peters, J., Janzing, D., & Schölkopf, B. (2017) *The elements of causal inference: Foundations and learning algorithms*. Cambridge, MA: The MIT Press.

Pietsch, W. (no date) 'Big Data – The New Science of Complexity'.

Popper, K.R. (1959) *The logic of scientific discovery*. Oxford, England: Basic Books.

Portmess, L. and Tower, S. (2015) 'Data barns, ambient intelligence and cloud computing: the tacit epistemology and linguistic representation of Big Data', *Ethics and Information Technology*, 17(1), pp. 1–9.

Powers, T. M. (2017) Philosophy and Computing: Essays in epistemology, philosophy of mind, logic, and ethics. Springer.

Prensky, M. (2009) 'H. Sapiens Digital: From Digital Immigrants and Digital Natives to Digital Wisdom', p. 11.

Ratti, E. and López-Rubio, E. (2018) 'MECHANISTIC MODELS AND THE EXPLANATORY LIMITS OF MACHINE LEARNING', *Machine Learning*, p. 18.

Reichenbach, H. (1938) *Experience and Prediction*. Available at: https://philpapers.org/rec/REIEAP-2 (Accessed: 14 December 2020).

Ribeiro, M. T., Singh, S. and Guestrin, C. (2016) '"Why Should I Trust You?": Explaining the Predictions of Any Classifier', *arXiv:1602.04938 [cs, stat]*. Available at: http://arxiv.org/abs/1602.04938 (Accessed: 24 September 2020).

Rieder, G. and Simon, J. (2016) 'Datatrust: Or, the political quest for numerical evidence and the epistemologies of Big Data'.

Rudin, C. (2019) 'Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead', *Nature Machine Intelligence*, 1(5), pp. 206–215. doi: 10.1038/s42256-019-0048-x.

Samek, W., Wiegand, T. and Müller, K.-R. (2017) 'Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models', *arXiv:1708.08296 [cs, stat]*. Available at: http://arxiv.org/abs/1708.08296 (Accessed: 24 September 2020).

Schmidt, M. and Lipson, H. (2009) 'Distilling Free-Form Natural Laws from Experimental Data', *Science*, 324(5923), pp. 81–85. doi: 10.1126/science.1165893.

Schurz, S. (2019) *Hume's problem solved: The optimality of meta-induction*. Cambridge, MA: The MIT Press.

Scott, Z. (2018) *Data Science's Reproducibility Crisis*, *Medium*. Available at: https://towardsdatascience.com/data-sciences-reproducibility-crisis-b87792d88513 (Accessed: 25 September 2020).

Shiffrin, R. M. (2016) 'Drawing causal inference from Big Data', *Proceedings of the National Academy of Sciences*, 113(27), pp. 7308–7309. doi: 10.1073/pnas.1608845113.

Spirtes, P., Glymour, C., & Scheines, R. (2000) *Causation, prediction, and search*. Cambridge, MA: The MIT Press.

Steadman, I. (2013) 'Big data and the death of the theorist', *Wired UK*, 25 January. Available at: https://www.wired.co.uk/article/big-data-end-of-theory (Accessed: 17 December 2020).

Stupple, A., Singerman, D. and Celi, L. A. (2019) 'The reproducibility crisis in the age of digital medicine', *npj Digital Medicine*, 2(1), pp. 1–3. doi: 10.1038/s41746-019-0079-z.

Symons, J. and Alvarado, R. (2016) 'Can we trust Big Data? Applying philosophy of science to software', *Big Data and Society*, 3(2).

Tsamados, A. *et al.* (2020) *The Ethics of Algorithms: Key Problems and Solutions*. SSRN Scholarly Paper ID 3662302. Rochester, NY: Social Science Research Network. doi:10.2139/ssrn.3662302.

Tukey, J. W. (1962) 'The Future of Data Analysis', in. doi: 10.1214/aoms/1177704711.

Turilli, M. and Floridi, L. (2009) 'The ethics of information transparency', *Ethics and Information Technology*, 11(2), pp. 105–112. doi: 10.1007/s10676-009-9187-9.

Waltz, D. and Buchanan, B. G. (2009) 'Automating Science', *Science*, 324(5923), pp. 43–44. doi: 10.1126/science.1172781.

Watson, D. S. and Floridi, L. (2020a) 'The explanation game: a formal framework for interpretable machine learning', *Synthese*. doi: 10.1007/s11229-020-02629-9.

Wheeler, G. (2016) 'Machine Epistemology and Big Data', *Routledge Companion to Philosophy of Social Science. Routledge* (2016).

Wigner, E.P. (1960) 'The unreasonable effectiveness of mathematics in the natural sciences. Richard Courant lecture in mathematical sciences delivered at New York University, May 11, 1959', *Communications on Pure and Applied Mathematics*, 13(1), pp. 1–14. doi:10.1002/cpa.3160130102.

Wu, C. F. J. (1997) 'datascience.pdf'.

Zednik, C. (forthcoming) 'Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence', *Philosophy and Technology*, pp. 1–24. doi: 10.1007/s13347-019-00382-7.

Zerilli, J. *et al.* (2019) 'Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?', *Philosophy & Technology*, 32(4), pp. 661–683. doi: 10.1007/s13347-018-0330-6.