VICI-project "Data science in psychopathology: gold rush in the data mine"

Jan-Willem Romeijn
Faculty of Philosophy
University of Groningen
j.w.romeijn@rug.nl

Summary of research project

New data science methods are changing research in psychopathology: clinical psychiatrists use automated classification methods to find subtypes of disorders, psychologists construct individual network models to unravel the causal structure of psychiatric disorders, and geneticists use "big data" methods to find their biological markers.

The data science methods hold the promise of data-driven and theory-free results. But a closer look at data science reveals that the theoretical suppositions are merely hidden from view. This raises concerns over the reliability and accountability of the research that is based on them. To address these concerns over the responsible use of data science, we need to uncover the suppositions inherent to them and understand their relation to the specific context of application.

The proposed research aims at such a clarification of data science methods. It employs a combination of formally oriented philosophy of induction, and practice-oriented philosophy of psychopathology. With this innovative combination of perspectives, we can illuminate data science methods conceptually as well as in their concrete scientific use. The result of this is a practice-oriented and formally precise epistemology of data science that can extend beyond psychopathology.

Three PhD projects investigate data science methods in three inter-connected research domains: data, models, and theory. Each project will draw out suppositions in the data science applications, combining formal and context-specific expertise. Two postdoc projects serve to share insights between the PhD projects, and connect them to the philosophy of induction and general philosophy of science.

The project offers a conceptual grip on data science methods in psychopathology, a better understanding of their general conditions of applicability, and improvements in their context-specific use and interpretation. Considering the rapid uptake of data science methods in social and medical science, and their impact on policy making and medical care, this is an urgent and important concern.

Keywords

Philosophy of science, data science, statistical inference, psychopathology.

Press release

Psychiatry and psychology make increasing use of data science methods. This project investigates these methods by using insights from the philosophy of science. The research is innovative in that it combines a mathematical understanding of the methods with knowledge of a scientific context in which these methods are applied, namely psychopathology. The applications of these methods are thereby illuminated, so that they become more reliable and more accountable in both the clinic and

the lab. The result is a practice-oriented and formally precise epistemology of data science methods that supports a better use of data science methods in the sciences.

Description of the project

Scientific relevance and challenges

Fool's gold?

Data science methods, comprising machine learning and other data-intensive or "big data" tools, are rapidly transforming a broad range of sciences. Cell biologists employ automated causal search to analyse the workings of cells, astronomers employ big data techniques to label stars in distant galaxies, climatologists estimate their models by ever larger data sets, and linguists make use of deep neural nets to improve automatic translators. In psychopathology, data science methods abound: automated clustering, feature learning and other data science methods help identify sub-types of mental disorders, support treatment decisions and predict courses of illness (e.g., Chekroud 2016, Aafjes 2021, Wang 2021, Wardenaar 2021).

There are high hopes for the gains of data science, and rightly so. Especially for sciences that target complex systems described by large numbers of interrelated variables, data science methods hold enormous promise. Traditional scientific methods like experimentation and statistical modelling struggle with such systems: how to construct the models when theory is by-and-large lacking, and how to structure and process the high-dimensional data? Machine learning and big data seem to provide convincing answers. They are famously heralded as the "end of theory" (Anderson 2008) and the "fourth paradigm" (Hey et al 2012), and there is even talk of a big data "revolution" (Mayer-Schönberger and Cukier 2013). A common theme in the expressions of optimism is that data science methods can make science genuinely data-driven.

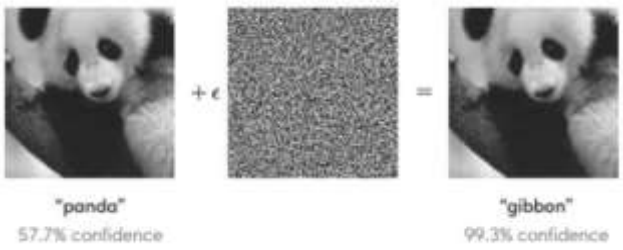
But beware! Fool's gold abounds in the data mines. Data mining may offer a way forward for sciences that target causally complex systems, but it may just as well lead us astray. Automated disease classification in psychiatry offers a disconcerting example. A pilot study (Beijers et al 2020), in which the PI joined a team of psychiatrists and statisticians, revealed that the application of automated clustering methods to a data set of psychiatric patients fails to produce robust patient groupings: the results of the clustering methods depend crucially on the parameter settings. Even on simulated data that was constructed with an embedded true clustering, variation of the parameter settings translated into wild variations over the clustering output. Not the data, but the method's parameters determine the labelling of patients. It is far from clear that data-driven clustering will facilitate reliable predictions of treatment response. Meanwhile, such methods are well-underway to impact clinical contexts (Kan et al 20XX).

This project ultimately aims to improve on the use of data science methods. Its immediate objective is to analyse their use in psychopathology research, focusing on the assumptions that underpin them, and to determine when they deliver domain-specific knowledge. It thereby offers an epistemological analysis of data science methods. The project contributes to the philosophy of science by bringing data science into view, and to the empirical sciences by offering tools to assess if data science methods are used responsibly.

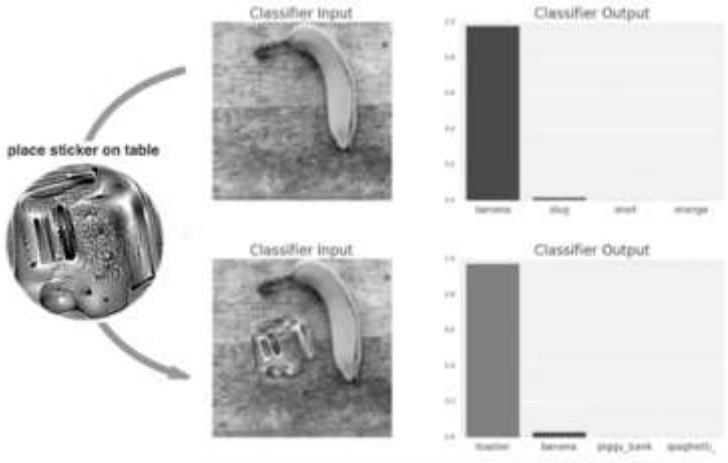
Black-boxed methods

Among AI researchers and computer scientists it is broadly recognized that the methods can misfire. However, to most scientific users the data science methods are “black boxed”: they cannot see how the methods work, and are therefore unable to recognize when they fail, resulting in a reliability problem.

So-called adversarials (Nguyen 2015) provide a telling illustration. A machine learning algorithm might correctly classify an image of a panda, as in the figure below. Stunningly, adding a reverse-engineered layer of noise to the image, thereby creating a marginally fuzzier but perfectly recognizable panda, will trick the algorithm into the inexplicable misclassification of a gibbon.



This problem generalizes. Mistakes of computers can be exploited systematically, up to the point where a banana can be made to look like a toaster by putting a sticker beside it that for the classifier evokes ultimate toaster-dom.



Admittedly, any predictive system is liable to occasional failures: adversarials for human vision have also been constructed (Elsayed 2016). However, the latter cannot be produced systematically and, by zooming out and reflecting, human beings are better able to control for their mistakes.

While image recognition continuously improves, the fundamental problem illustrated by adversarials remains: data science methods are opaque. It is often difficult to trace mishaps back to their causes, and therefore difficult to repair the methods or hedge against mistakes. It is for this reason that, at the conference of the Neural Information Processing Society in 2017, Ali Rahimi termed machine learning the “new alchemy” and called for a “rigor police”, i.e., for more research into

conceptual and mathematical controls on the methods (Rahimi and Recht 2017, Shalev-Shwartz 2019, LeCun 2019). Within computer science and AI, such research efforts are ongoing.

Summing up: there are reasons to worry over the reliability of data science methods, far beyond the use of data science methods in psychopathology. Part of the response to these worries may be a better technical understanding of the methods, a task mostly falling to data scientists themselves. But as argued below, scientific users need an understanding of how the methods, as embedded in scientific practice, can deliver knowledge. This latter task is taken up in the present proposal.

Accountability of data science

Concerns over the reliability of data science methods are matched by similar concerns over their accountability. In her book "Weapons of Math Destruction" (2016), former data analyst and big data expert Cathy O'Neil writes:

> "Here we see that models, despite their reputation for impartiality, reflect goals and ideology… Models are opinions embedded in mathematics."

The models referred to here are based on machine learning over large databases. O'Neil's point is that these data science methods are charged with assumptions of a political and moral nature (cf. Eubanks 2018, Barocas and Selbst 2016). To uncover the ideological goals and commitments that reside inside data science methods, they are in need of further unpacking. Concerns of this kind have over the last years lead to a wave of research into the ethics of data science, and into so-called explainable artificial intelligence, or XAI for short (cf. ACM FAccT 2021).

Another promoting cause for ethical interest in the transparency of data science is the development of EU laws entitling citizens to an explanation of decisions based on artificial intelligence (cf. Doshi-Velez and Kortz 2017, Goodman and Flaxman 2017). The use of data science methods runs into trouble here, as decisions based on these methods are mostly not supplemented by explanations. Closer to the Dutch context, the national science agenda (Nationale Wetenschapsagenda 2017) recognizes responsible big data research as one of its core themes, and includes transparency of data science methods among its key objectives.

The dangers of opaque data science methods are illustrated in ProPublica's by now classic investigation into a data-driven system that determines risk scores for recidivism of convicts in the US judiciary (Angwin et al 2016). The system turned out to be discriminatory: for groups with higher chances of conviction the error rates are higher, thereby placing groups that are already vulnerable at a disadvantage. Such statistical biases are known from psychometrics (cf. Borsboom, Romeijn and Wicherts 2008), and in the debate over fairness in AI they were effectively rediscovered (Kleinberg et al 2017).

In sum, data science methods need to be made more transparent in the interest of accountability. Once we base major policy decisions on them, it is on us to motivate and explain ourselves. This ethical motivation for transparency dovetails with the epistemic motivation of this proposal: responsible data science requires both. XAI research presents an important step in this direction. But it often focuses on the ethical and political dimensions of data science, rather than on the epistemology, and it does not take into account the specifics of the scientific contexts in which the methods are applied. Such an epistemology for the scientific use of data science is taken up in the current proposal.

The case of psychopathology

The problems of data science are general, but not all sciences are exposed to these problems to the same degree. Psychopathology is arguably among the sciences that are most badly affected: its domain of investigation is large and causally complex, ranging over many scales and levels of description, from the genetic all the way to the socio-economic (Kendler et al 2020). The space of possible predictors and causal relations is correspondingly vast. Moreover, in comparison to other causally complex sciences there is relatively little consensus over the theoretical and mechanistic knowledge that can help data science on its way in determining predictors and causes (cf. Ross 20XX, Poland and Tekin 2017, Fried 2020).

To explain this further, consider planetary astronomy as a base case. The systems under scrutiny in this science are described by few variables, and as a result the space within which we can search for relevant empirical patterns is limited. As nicely illustrated in Thagard (1993), a basic search algorithm can identify Kepler's third law of planetary motion. Next consider cell biology, the science that studies the myriad chemical processes in the cell that together constitute gene expression, protein folding, and so on. Cell biologists are confronted with a large space of possible predictors and causal pathways, and this complicates the application of traditional scientific methods. In such a causally complex science, data science methods may be used to great effect (Mooij 2020). However, in this context the use of data science is crucially supported by mechanistic knowledge about processes in the cell, which allows us to impose constraints on the search space and help automated discovery on its way.

Now contrast this with psychopathology. While its complexity arguably surpasses that of cell biology, there is very little psychopathological theory to constrain the machine learning and big data research on it. The automatic clustering towards causally homogeneous mental disorders, for instance, cannot rely on any basic physiological or genetic characteristics that patients with the same disorder will always share. It is therefore harder to ensure that the clustering method latches on to something relevant or useful. Similarly, in genome-wide association studies (GWAS) and other big data efforts in psychopathology, the methods used do not get any steering from facts about what combinations of variables may be relevant. Instead, the data are simply mined for mostly tiny correlations that are robustly present, and these correlations are best taken as suggestions for further study, not as providing proof for the genetic basis of mental disorder (Matthews and Turkheimer 2021). The point is that, because psychopathology has fewer theoretical constraints to rely on, it will be more vulnerable to the methodological mishaps described above and the concomitant problems of reliability and accountability.

Sciences differ in how vulnerable they are for methodological mishaps, but also in their potential for causing harm. Misclassifying a star or galaxy is unfortunate, but it is not as ethically problematic as the misclassification of a patient, because the latter mistake will directly influence interventions in the clinic. Concerns over reliability are more pressing for sciences whose results impact on our lives, psychopathology among them. And the same goes for concerns over accountability. For example, psychiatrists analyse the smart-phone data of patients to predict the course of illness (e.g., George 2021). But to communicate with patients and explain therapeutic interventions, e.g., keeping a patient under supervision based on a warning signal, they will need an explanatory story, over and above the data-scientific prediction results. It will negatively affect the trust and understanding of patients if doctors rely primarily on intransparent data-science.

For all these reasons, psychopathology is well suited as case study into the use of data science methods. The ingredients are all there: it is causally complex and mostly lacks a shared theoretical framing, and the problems of reliability and accountability are immediate. Moreover, as further argued below, psychopathology can stand in for a much wider range of social and medical sciences.

Innovation

A context-sensitive epistemology of data science

Data science holds a promise for all sciences that face high-dimensional data. However, it can only live up to that promise if we manage to address the reliability and transparency problems described above. And to address these problems we are in urgent need of an epistemology of data science.

The key idea of this project is that the philosophy of science can provide this epistemology by a two-pronged approach: a formal and conceptual analysis of induction, in combination with a practice-oriented philosophy of science. The project ventures into the new domain of data science, an area that is under-researched in the philosophy of science, pairing a detailed mathematical understanding of data science to a deep engagement with a specific application domain: psychopathology. This combination is unique. Philosophical researchers who have sufficient mathematical background for grasping the data science do not normally have access to a relevant application domain like psychopathology, and philosophers specializing on such an application domain often lack the mathematical expertise for analysing the methods in detail.

It deserves emphasis that the combination of these approaches is essential for providing real assistance to psychopathology. There is a rich formal literature in the philosophy of induction and statistics to which the PI actively contributes, but this literature is not generally accessible to methodologists working in psychology and psychiatry. And there are numerous insights from general philosophy of science that can prove their value to psychopathological research, witnessing extensive work of the PI on psychopathological methods, classification tools, and research practice. To engage with the new topic of data science, and effect real change in how data science methods are used and evaluated in psychopathology, we need an integrative approach that addresses both the technical detail and the context-specific significance and interpretation of the data science methods.

Importantly, the successful execution of this project will achieve much more than a critical appraisal of data science methods in psychopathology. The themes and concerns described above are general, and they show up in a wide variety of social and medical sciences, or “human sciences” for short (cf. PLoS Medicine Editors 2018, Ching et al 2018). Many human sciences are confronted with target systems that are forbiddingly complex, e.g., the brain, cognition, social interaction, cultural conflict, and so on. In response to this, the human sciences have responded by adopting a statistical methodology, and this has arguably curbed their theoretical development (cf. Eronen and Romeijn 2021). For these “statistified” human sciences, the step to data science methods is a natural expansion of existing methodology. However, partly for their statistical methodology, these sciences are more exposed to the problems of reliability and accountability outlined in the foregoing. The insights from this project will therefore be relevant to a whole range of such human sciences.

The challenges of data science are substantial, but it bears repeating that the promise of data science outweighs the challenges. The “gold rush in the data mines” is there for a reason. The eventual goal of this project is to help data science unearth the knowledge that it promises. Looking down

history we see that philosophical works have often supported scientific and methodological developments: Bacon and Boyle paved the way for the experimental method, the works of Poincaré and Mach were important for modern physics, and Popper's *Logik der Forschung* (1934) contributed to a methodology for the social sciences. The development of statistics has been accompanied by several philosophical works that promoted the adequate use of the statistical methods (e.g., Fisher 1956). This project springs from a similar motivation: to contribute to the responsible introduction of data science methods within a range of human sciences, and to prepare these sciences for a new data science era.

Positioning the project

This project does not venture into entirely uncharted territory. Data scientists see the need for methodological guidelines (Breiman 2001, Hey et al 2012, Diggle 2015, Lipton 2016, Caruana 2017, Kuffner and Young 2017, Roscher et al 2020), based on internal arguments concerning the adequacy of data science methods. Insightful descriptions of these methods, mostly focusing on computational and mathematical detail, are available in data science itself, and efforts to clarify them are ongoing in the XAI movement. However, in the data science community the use of data science methods as part of a broader scientific method has not seen much targeted attention. And understandably so: it does not fall within the purview of data science itself to integrate its methods into scientific methodology.

As briefly discussed above, there is already an active research area occupying the intersection of philosophy and data science, relating to ethics and explainable AI, and pertaining to the accountability of data science. The current project, by contrast, is epistemological in nature. It thereby opens up a new and mostly unexplored direction for research on the intersection of philosophy and data science. Moreover, epistemological insights into how data science methods produce knowledge, into how and when they can be applied successfully, are arguably a pre-condition for any full account of responsible science. The project lays the groundwork for any further debate along ethical or political lines.

Notably, the scientific use of big data methods has been targeted more extensively in science and technology studies (e.g., Beaulieu 2004, Leonelli 2016, Mackenzie 2017). This literature focuses on how institutional contexts, research cultures and disciplinary boundaries impact on the practice of data science, offering invaluable insights into the relevant scientific practices. However, it does not engage with the mathematical details of data science methods. It does not study these methods in their formal epistemological capacity, as attempts to make justifiable inductive inferences.

The current project differs from these existing research traditions on philosophy and data science. It targets the use of data science methods in their role of producing scientific knowledge, against the backdrop of both general and context-specific philosophical expertise, thereby differing from data science, XAI, and ethical approaches. Furthermore, it engages with the mathematical and computational details of data science, viewing the methods as tools for inductive inference, thereby differing from the more sociological work done in science studies. Despite calls for action (Williamson 2004, Corfield 2010, Ortner and Leitgeb 2011, Romeijn 2014a, Kitchin 2014, Wheeler 2016, Leonelli 2016), and with few notable exceptions (e.g., Creel 2020), the proposed research is unique in combining formal and practice-oriented philosophy of science. It fills a large gap in the literature, by providing a technically precise epistemology of data science that is grounded in scientific practice.

Methods

Approach

The insight that the data do not speak for themselves is a venerable one. Francis Bacon (1620/2014) already warned against the unreflective collection of data, likening empirical scientists to ants who "heap up and use their store" and recommending the approach of the bee that "extracts matter from the flowers of the garden and the field, but works and fashions it by its own efforts". The mindset of this project is very much along these lines. The pretence of data science methods is that they rely predominantly on data. But scientists always mix some theory into their inferences from data, through data construction, model-based inferences, and theory-laden interpretations of the results.

To apply data science responsibly, the theoretical suppositions inherent to data, models and theory need to be revealed, giving the users of data science a better insight into its conditions of applicability. As said, this overall approach is worked out in two interrelated philosophical sub-disciplines.

1. Formal philosophy of science. The focus on this side of the project is on the philosophy of induction. Mathematical tools, like probability theory and inductive logic, can help us to unravel and assess scientific inference. The project relies on these tools to offer a conceptual clarification of data science.
2. Philosophy of science in practice. Here the focus is on a practice-oriented and socially engaged philosophy of psychopathology. Modern philosophy of science often works in close contiguity with the sciences it studies, employing the interpretation and conceptual analysis of theories and models as its method.

1. Formal philosophy of science

How can we bring the implicit suppositions of data science methods to light? Conceptual resources for this can be found in formal philosophy of science, specifically in inductive logic and the philosophy of statistics. A central theme in these disciplines is to uncover assumptions in the process that leads from data construction, through inference, to predictions and decisions. The strategies for doing this can be transferred onto the analysis of data science methods.

Adversarials revisited

Before turning to inductive logic and philosophy of statistics, it is helpful to discuss a general theme from these disciplines (cf. Howson 2000, Romeijn 2005). It is the idea, already voiced by Bacon, that nothing can be learnt from data alone, unless those data are framed and processed in a particular way, and thereby provided with theoretical content. The ideal of an elimination of theory from induction is a recurrent one (cf. Hartmann 2011, Romeijn 2014a) but as data scientists will readily admit, theoretical suppositions are just as inevitable in data science as elsewhere.

This insight is nicely illustrated by a debate on inductive logic. Carnap (1950, 1952) aimed to justify predictions on the basis of data alone by relying on a notion of logical probability. He hoped to realize the ideal of theory-free inductive method, and arrive at an empiricist foundation for scientific knowledge. Putnam (1963) challenged the program, pointing out that if Carnap was correct, "science could in principle be done by an idiot", i.e., by a machine carrying out elementary computing instructions. Putnam drove his point home in a way that is highly informative for current data science.

He defined sequences of data that, for a given inductive logic, are fundamentally unlearnable, thereby showing that any presumably theory-free inductive method has an Achilles' heel. Such a method will misfire when it is confronted with data that are deliberately constructed to be at variance with the theoretical suppositions implicit in the method.

It is remarkable how the construction of Putnam foreshadows the so-called adversarials of machine learning that were depicted in the foregoing. We can always construct adversarial data sequences that throw sand into any inductive learning machine. Moreover, the lessons drawn from this possibility of adversarial data are similar. Adversarials reveal that there are theoretical suppositions, or "inductive biases", inherent in any inductive method. This bias makes the inductive methods suitable for application in one context, but unsuitable in another. It reveals, in other words, that all such methods have conditions of applicability. Putnam's challenge to Carnap shows how an insight that has been around in inductive logic for decades can help us to understand and re-evaluate current problems with the reliability of machine learning.

Inductive logic

There is a general similarity between machine learning methods and Carnapian inductive logic. A striking historical link between the two disciplines is presented in the work of Solomonoff. As expounded in Sterkenburg (2018), Solomonoff (1964) was inspired by Carnap's work on universal induction, and it subsequently served as an inspiration to the machine learning community (e.g., Hutter 2007). The idea of a purely data-driven inductive method thus finds one of its intellectual roots in inductive logic. Moreover, the historical link is reinforced by a conceptual one: inductive logic and machine learning both claim to be based on given data only, and both venture to provide reliable predictions of future data on that basis. For Carnap and followers, a further basis for these predictions was found in the concept of logical probability (Zabell 2012). For data science, the further basis for the data-driven methods is not quite so clear.

In this research proposal the prediction rules of inductive logic are viewed as a proto-version of the methods of data science. Inductive logic houses a long and rich tradition of conceptual research on prediction, for instance on the notion of logical probability (Hartmann 2011). Some of this research focuses on axiomatic foundations of predictive systems (Paris and Vencovská 2015, Huttegger 2017). Other such research bridges the gap between inductive logic and statistical inference (e.g., Romeijn 2011).

It is well-known that the prediction rules of inductive logic can be translated into Bayesian statistical inferences (De Finetti 1937, Skyrms 1996) via De Finetti's representation theorem. This theorem, like its generalizations, shows a correspondence between data-driven predictive systems on the one hand, and Bayesian statistical methods on the other. The predictive systems are thereby given a redescription involving a statistical model, which allows us to identify the theoretical suppositions that otherwise remain implicit in the predictions (Romeijn 2004). Such redescriptions provide a blueprint for the conceptual analysis of data science methods. Any data science method can be converted into a corresponding model-based statistical system, through the development of a tailor-made representation theorem. That this is possible can be seen in Romeijn (2006), Paris and Vencovská (2015), and also more recently, in Sterkenburg's (2018) thesis, which was co-supervised by the PI.

Another insight from inductive logic plays a similar role in uncovering implicit suppositions: the conceptual basis for predictions is inherent in the language with which the observations are described (cf. Sznajder 2017). This insight suggests that we can analyse data science by focusing on the construction and preparation of the data themselves. Many of the theoretical suppositions inherent in data science methods can be traced back to the data construction. A good example is the use of a specific metric over a space of psychological attributes, which determines similarities and dissimilarities of patients located in that space, and thereby steers the automated clustering method.

Philosophy of statistics

Besides inductive logic, the project deploys insights from the philosophy of statistics. Many statistical procedures rest on the empiricist ideal that the data are the sole basis for the assessment of statistical hypotheses. This ideal chimes with the ambitions of data science and indeed with a broader scientific ideal of objectivity (cf. Daston and Galison 2007). But also in statistics, it is apparent that theory-free methods are unattainable.

We can perceive the ideal of theory-free science in much of the foundational literature on statistical testing, estimation (Neyman and Pearson 1967, Fisher 1956, Barnett 1999) and fiducial inference (Seidenfeld 1979, 1992). All these statistical procedures are defined as functions over the sample space, taking only data as input, and all of them output a verdict on a hypothesis, or else select a hypothesis from a range. That tests and estimations are merely functions over sample space supports their portrayal as data-driven methods. Additionally, many of the methods for the evaluation of statistical models are similarly data-driven. This obviously holds for methods based on cross-validation (Hastie et al 2001), but it is also true for several of the information criteria, e.g., AIC and its relatives (Claeskens and Hjort 2008).

Analyses from the philosophy of statistics show that such methods invariably rest on implicit suppositions. Many of them take the form of assumptions about the statistical model on which the test, estimation, or model evaluation is based, or else they take the form of a prior probability over the model. For example, fiducial probabilities rely on a functional model (Dawid and Stone 1982), which details particular dependence relations between stochastic and systematic components of the target process. Stein's shrinkage estimators, used in the meta-analytic aggregation of statistical studies, are effectively resting on an empirical prior (Efron and Morris 1973). And model selection methods carry implicit suppositions on what penalty for complexity is adequate, which in turn rely on assumptions about the distribution over data and the measure of divergence between distributions (Kieseppä 1998 and 2003, Romeijn 2017).

Further suppositions in statistical procedures can be located in the construction and framing of the data. A clear illustration of this is provided by the so-called stopping rule controversy (e.g., Sprenger 2009, Steele 2013). Many classical statistical procedures depend not only on the data that is actually obtained, but on the shape of the space of all possible data, i.e., the sample space, which is in part determined by rules on when to stop collecting data. By some this is considered a violation of the idea that science is based on empirical fact only (cf. Birnbaum 1962, Berger and Wolpert 1988), while others maintain that the dependency of statistical procedures on what could have been observed, in addition to what actually was observed, is in a sense correct (de Heide and Grünwald 2018). Whatever one's position on this, it reveals that statistical procedures import theoretical assumptions through the way in which data are framed.

Much of the philosophy of statistics is devoted to uncovering suppositions of this kind, and thereby clarifying the procedures. The strategies for uncovering these suppositions can be carried over to applications of data science. A quick example is offered by causal modelling methods proposed to reconceptualise mental disorders (Borsboom and Cramer 2013): statistically the factor-analytic and network models are equivalent but these approaches are different owing to implicit theoretical suppositions (van Bork 2019). A conceptual analysis of the approaches can help tell them apart.

2. Philosophy of science in practice

In the foregoing I discussed the two intellectual resources for the current proposal: inductive logic and the philosophy of statistics. The project uses these resources to study data construction, inference from data, and the deployment of findings, as they manifest in specific applications. Data science methods are thus considered in technical detail, and mostly on a conceptual level. But to make the results of these analyses relevant and convert them into actionable advice, we need to supplement the formal approach with an orientation on concrete practices.

Focus on psychopathology

As already argued, the research area of psychopathology presents an optimal case-study. It stretches the social and the medical realm, it does not have a unified and established theoretical framework, its results are directly relevant to people's lives, and data science methods are rapidly transforming its methodology.

Besides the appropriateness of psychopathology as a case study, the philosophy of psychopathology provides numerous points of connection for the project. Alternatives and revisions of the classification of disorders, like HiTOP and DSM5, are hotly debated, as are methodological criteria for justifying such revisions and philosophical interpretations of the classifications themselves (van Loo and Romeijn 2018). Efforts to connect statistical and technical analyses to interpretative ones have already been undertaken, and indeed help to resolve concrete methodological challenges, like comorbidity and nosological reform (e.g., van Loo and Romeijn 2015, 2019, 2020). Moreover, there is wide-spread recognition within philosophy of psychopathology that its epistemological problems cannot be considered in isolation from societal impact (Cooper 2014, Schaffner and Tabb 2015, Tabb 2019). The challenges that follow from the introduction of data science in psychopathology can thus be connected very easily to a lively research area.

Several debates in the philosophy of psychopathology relate specifically to the uptake of data science methods. Under the header of precision medicine there is growing interest in tailoring medical care to the individual case, and therefore in methods that allow us to track and predict on an individual basis (cf. Juengst et al 2016). Machine learning on personalized data is naturally seen as a way forward. However, stylized findings in a database do not automatically translate to interventions in the clinic. Practical deployment requires reliable predictions, but also an explanation of these predictions by reference to an interpretable and transparent model. Absence of such explanations hampers the acceptance of the medical interventions by the patients, and it also obscures questions of responsibility (Nevin 2018).

Another important debate within psychopathology concerns the intricate relation among its numerous levels of description, scales, and domains. Data scientists may take an "a-reductionist" stance on whether any scale or domain takes priority, and employ variables for predictive purposes

irrespectively of where they can be located (cf. Romeijn and van Loo 2020). On the other hand, there is pragmatic value in adhering to a theoretically motivated vocabulary that can facilitate translations across research domains (cf. Kendler 2012, Kendler et al 2020).

Relatedly, philosophers and psychologists have started to lament the lack of psychopathological theory (Fried 2020), arguing that the availability of theoretical structure will help direct research efforts in psychopathology, and ultimately improve our ability to predict and intervene. The call for a better grasp of the causal structure of mental illness is a good example of this (Borsboom and Cramer 2013). The speedy uptake of data science methods is arguably a cause for concern here, precisely because the theoretical commitments of these methods are not readily accessible. The theoretical development of psychopathology is curbed by the use of data science methods.

In sum, the philosophy of psychopathology offers numerous promising inroads for investigating and clarifying the problems of data science, precisely where these methods are applied. Research themes include the construction of data that serves as input to the data science methods, the interpretation of classification systems for mental illness, the relation between empirical and theoretical aspects of such systems, and the nature and justification of clinical interventions. It is of vital importance for the relevance of a mathematical and computational understanding of data science methods that we also understand their role in the concrete practice of a science, and the philosophy of psychopathology presents a wealth of opportunity for this.

Practically oriented and socially engaged

Good philosophy of science is often done in collaboration with scientists, by a method of participant observation. The proposed research adopts this method whole-heartedly. An important advantage of this proximity to practice is that the philosophical insights can also be used to improve the applications of the scientific findings in policy making (cf. Cartwright and Hardie 2012). This signals another important orientation of the current project, towards promoting responsible applications of science and engaging with questions that have societal relevance.

Recall that social engagement is an important motivation for focusing on psychopathology. Findings within this discipline often have implications for our lives, and the afore-mentioned problems for data science are therefore particularly pressing for them. Philosophical research into these problems ultimately serves a higher goal: transparent, reliable, accountable, and hence responsible science. It supports a vision on society and science in which the latter is a facilitator of democratic empowerment and a motor of social change (e.g., Longino 1990, Kitcher 2003). As such this project stands in a long philosophical research tradition that has its roots in the Enlightenment, and is aimed at clarifying the scientific enterprise. This tradition encompasses logical empiricists like Neurath, Carnap and Hempel, but also the work of Kuhn (1962), Latour (1987), van Fraassen (1980) and Douglas (2009).

In this Enlightenment spirit, the project offers clarity on what data science methods can do for us, and a set of rules that can guide us in responsible applications. In its criticism and improvement of data science, the project exhibits a particular stance on the role of philosophy of science vis-à-vis the sciences: it focuses on concrete practices in the sciences, and positions these sciences in a wider context that has ethical and societal dimensions. That is, it promotes a socially engaged philosophy of science in practice, equipping science with better tools for serving its societal role.

Project setup

Project basics

Subprojects at a glance

The foregoing identified three domains of investigation for an epistemology of data science: data, models, and theory. The three PhD projects below are associated with these domains.

1. Big data: the emphasis of this project is on the framing and processing of data, and the theoretical assumptions that are thereby imported. Central case studies involve the analysis of high-dimensional behavioural data by means of multiple regressions, and the nature of genomic data.
2. Machine learning: this project analyses specific machine learning methods, attempting to uncover the modelling assumptions in them. Here the focus is on disease classification using automated clustering methods, and on the analysis of time series with idiographic data.
3. Missing theory: this project is concerned with causal modelling and network methods in psychopathology. It develops an account of how causal theory facilitates clinical research. The overall theme is the role of theory and how it helps reconceptualising psychopathological phenomena.

Two postdocs investigate the three domains of data science integrally, from the perspectives of philosophy of science in practice and mathematical philosophy respectively. The subproject of the PI covers both. All three support the PhD students in the early stages of their projects and build bridges between them.

4. Observations and models in data science: this sub-project scales up from the project's case studies towards lessons for the philosophy of science in general.
5. Inductive inference in data science: this sub-project connects the project's case studies to methodological discussions within data science and statistics.
6. The PI integrates the sub-projects and thereby offers an epistemological clarification of data science, in psychopathology and beyond.

HR

The ideal PhD candidates have a background in philosophy of science and in the methodology of psychology or psychiatry, e.g., psychometrics or epidemiology. The postdocs will be recruited among recent philosophy of science PhDs. Familiarity of the researchers with the practice of data science is crucial to the project's eventual value.

Collaborations

The PI has established working relations with scientists from all disciplines involved. He is active in an expert network interfacing philosophy of science, statistics, and machine learning, and similarly active in a network of psychiatrists, psychologists, and philosophers of psychopathology. This network includes colleagues from the Netherlands and from high-ranking universities in Europe (LSE, Bristol, LMU Munich, Düsseldorf, Paris, and Turin), Australasia (ANU, Fudan Shanghai) and Northern America

(Pittsburgh, Carnegie Mellon, Caltech, UC Irvine, and Maryland). Members of the project's advisory board are:

- Prof. F. Eberhardt, Caltech
- Prof. S. Huttegger, UC Irvine
- Prof. K. Kendler, Virginia Commonwealth University
- Prof. H. Leitgeb, Ludwig Maximilians University
- Prof. S. Leonelli, University of Exeter
- Prof. M. Solomon, Temple University

PhD students will be stimulated to visit board members as guest researchers in their second or third year. Members of the advisory board will be invited to speak at the four workshops.

Planning and deliverables

The table summarizes the planned project output by half-years, listing journal publications (P), theses (T), and a book (B). The project team also delivers four focused workshops (W), a set of short films (F), and a guidelines document for empirical researchers (G). Further details on the planned publications are provided in the detailed descriptions of the subprojects below.

Role \ Period	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b	7a	7b
PI			P		P		P		P		P		B	
PhD				P		P	P		T					
PhD				P		P	P		T					
PhD						P		P	P		T			
Postdoc			P	P	P	P								
Postdoc					P	P	P	P						
Project				W		W		W	F	W		G		

The guidelines are aimed at helping scientists to apply data science methods more responsibly, and interpret their results in the right way. Journals and professional organizations issue such guidelines for reporting statistical findings (cf. Wasserstein and Lazar 2016, Open Science 2021). The guidelines serve a similar goal of promoting transparency.

<u>Overview of subprojects</u>

Project 1 "Big data"

The emphasis of this sub-project is on the identification of inductive suppositions that are embedded in data structures. It investigates how the construction of data sets imports such assumptions into the research.

- A contender for mental disease classification is HiTOP (Kotov 2017). Its general idea stems from psychometrics: a large number of psychological variables is analysed by means of exploratory factor analysis alongside PCA and other data reduction techniques, leading to a seemingly data-driven, dimensional classification system. What remains mostly invisible in the identification of salient constructs is that the manifest variables themselves, representing

symptoms and psychological attributes, are charged with theoretical assumptions. Choices on how to operationalize or aggregate such variables matter greatly to the end result (Wilshire et al 2021). The process of importing suppositions in the data construction can be made transparent by relying on a large literature, both technical and general, about the theory-ladenness of observations.

- Psychiatric genetics involves the analysis of statistical relations between genetic and behavioural variables in so-called genome-wide association studies (GWAS). It may seem that GWAS data, which are combined in polygene scores to indicate patients' risk for psychiatric disorders, are "theory-free". However, the way in which these data have been collected and labelled is inevitably steered by theoretical presuppositions about the diseases for which they are supposed to be relevant (cf. Leonelli 2016, 2020). The appeal to data being "theory-free" calls for philosophical scrutiny.

Notice that the themes from inductive logic and the philosophy of statistics, namely that theoretical suppositions are contained in the labelling of the observations and in the construction of the sample space, are here connected to concrete examples of psychopathological research data: theoretical suppositions are imported into the research by the way in which the data input is structured.

Project 2 "Models and machine learning"

Here the emphasis is on the identification of modelling assumptions in specific data science methods. The case studies are drawn from psychiatric disease classification. In recent years, researchers in psychiatry have started using data science methods to improve on their classification efforts (cf. Blaauw 2017, Wang 2021, Wardernaar 2021), complementing and sometimes replacing traditional statistical approaches. But these approaches often fail to take proper note of the underlying suppositions in the inference processes that lead to classifications (cf. Beijers et al 2019).

- The PI is on contact with a consortium of psychiatrists that is currently designing a data science system for predicting treatment response, based on automated clustering algorithms (Kan 20XX). Following the afore-mentioned equivalence of predictive systems and statistical hypotheses, we can reconstruct what statistical hypotheses are driving the predictions, and check these against the contexts in which they are applied.
- Big data psychiatry employs so-called Lasso and Ridge regressions on behavioural data sets, involving the aggregation of repeated multi-variate linear regressions to find natural patient groups (Kessler 2016). These techniques help researchers to select variables for inclusion in a predictive model, and hence they are close to model selection methods. A comparison between these two types of analysis will offer insight into the implicit modelling assumptions that drive the former.

Once again, themes from formal philosophy and from the philosophy of science in practice can be recognized in these concrete applications of data science methods in psychiatry. We are looking for representations of machine learning methods in terms of statistical hypotheses, or for an understanding of automated model construction in terms of statistical model evaluation, all against the background of a clinical practice in which these classifications facilitate treatment interventions.

Project 3 "Missing theory"

The emphasis of this sub-project is on the use of psychological theory in data science, and the way in which new theoretical notions allow researchers to reconceptualise the phenomena, and thereby facilitate better predictions and interventions. A highly relevant development in this regard is the perceived "theory crisis" of psychology (Fried 2020): it collects and categorizes observable phenomena, but it lacks a theoretical basis, and therefore it fails to systematize the findings, let alone direct further research efforts.

- With the dominance of the Diagnostic and Statistical Manual of Mental Disorders (DSM), psychiatry grew increasingly operationalist. In recent years, however, there have been several attempts to deploy causal modelling methods into psychopathology, most notably in the network approach to psychopathology (Borsboom and Cramer 2013). Interestingly, this reintroduces theoretical, namely causal, structure into a predominantly empiricist domain. The subproject will investigate whether such models, designed first and foremost by computer scientists to support automated causal search (Spirtes et al 2001), can carry such theoretical interpretations, and whether these causal structures are advantageous for making predictions.
- Network models of mental disorder are increasingly popular in psychopathology. However, such models are statistically equivalent to specific latent variable models (van Bork 2019). The former arguably have computational and interpretative advantages, but they also introduce modeling possibilities that obscure the relation to the target system, especially in the area of network analysis (Bringmann 2019). This raises questions on the meaning of data science results based on network simulations.

As before, this project combines an orientation on the practice of psychopathology with a detailed study of how various data science methods, in particular causal modelling, function within the science.

Project 4 "Observations and models in data science"

This project provides an overview of how general philosophy of science, and in particular the philosophy of psychopathology, relates to the case studies and their project topics. The goal here is to disclose existing philosophical work on data, models and theory, and find points of contact between this literature and the data science methods that are studied in the PhD projects. Potential topics for further investigation are the following.

- While the philosophy of science offers a fairly well-developed theory on how to handle evidence, there is relatively little material on how data turn into evidence in the first place (cf. Morey et al 2016), let alone the specific evidential impact of large quantities of data. Both the process of data construction and the confrontation of data with a statistical model are involved in this. The philosophy of science literature on measurement holds valuable insights that can be put to work in the clarification of data science methods.
- Theoretical considerations can enter into statistical procedures in several ways: through the choice of a prior or other constraints on a model, through the specific shape of the sample space, and so on. Besides this, theory enters statistical methods via the context in which the methods are applied (cf. Leonelli 2016). Relying on the taxonomy of Creel (2020), similar entry points for theory can be identified in big data and machine learning methods.

An important goal of this sub-project is to offer the PhD students a good understanding of the way in which the domains of data, model and theory hang together, so that they can exchange insights among themselves.

Project 5 "Inductive inference in data science"

This project provides an overview of methodological discussions within data science itself, and relates them to the PhD projects. One important task is to assist the PhD candidates in coming to grips with the data science methods that they study. Another is to determine the relevant relations between data science on the one hand, and inductive logic and philosophy of statistics on the other. The project will be carried out against the backdrop of a large literature on inductive learning, in the philosophy of science (e.g., Kelly 1996, Schurz 2019) and in theoretical computer science (e.g., Vovk 1998, Schölkopf and Smola 2001, Vapnik and Kotz 2006, Harman and Kulkarni 2007).

- As discussed in the foregoing, there is a striking parallel between adversarial data in machine learning and the older criticism from Putnam on inductive logic. While there is an obvious similarity on the surface level, a more thorough investigation will likely reveal interesting disanalogies as well, especially on the nature of the inductive suppositions that expose the learning methods to adversarials.
- Neural networks are highly successful and versatile when it comes to picking up patterns in the data. This raises concerns over their potential to "overfit", i.e., to read a systematic component in what are merely perturbations, mistaking the noise for a signal. It is why practitioners sometimes refer to deep learning as "glorified curve fitting". A comparison between machine learning and other methods for constructing or selecting statistical models will illuminate if and how the former avoids the problem of overfitting. Probabilistic theories of neural networks (Patel et al 2015) provide a starting point for this research.

This sub-project opens up a wealth of technical insight for the PhD students, and will help them navigate a fast-growing literature on data science methods.

Project 6 "The epistemology of data science"

The primary aim of this sub-project is to write a monograph and draw up a guidelines document for scientists who work with data science methods. This is achieved by bringing the results of the other subprojects together, trying to find commonalities among them and developing a view on knowledge production through data science in which the domains of data, models and theory are integrated.

Outreach

<u>Mission</u>

Insight into the use of data science methods is both urgent and important. Considering the speedy deployment of data science methods and their direct impact on clinical decisions, the relevance for psychopathology will be apparent. But the relevance extends towards other so-called human sciences, and indeed to all research areas that use data science. Moreover, the same urgency and importance is apparent in societal contexts. For professionals in governments and businesses, responsible decision making requires an understanding of the results of data science. And for a wider public, as

underscored by the reliance on search engines and social media, the importance of providing access to, and public control over data science can hardly be overstated. The outreach activities of this project are therefore not an afterthought. They present an integral and crucial part of the project.

Relevance for psychopathology and beyond

The applicant has an excellent track record in collaborating with scientists from a wide range of disciplines. There are working relations with researchers from all disciplines involved. The format for the collaborations is one of mutual support, with scientists informing philosophers on their use of data science, and philosophers informing scientists on the foundations, conditions of applicability, and interpretations.

Research contacts

The project is connected to a large number of scientific contexts in which clarity on data science methods is needed.

- With dr. H. van Loo and prof. R. Schoevers (Psychiatry RUG) the applicant has worked extensively on classification methods in psychopathology, ensuring direct ties to the psychiatric clinic of the university. Highly relevant is the contact with dr. F. Jörg, who coordinates I-SHARED, a project that employs data science methods in the clinic.
- Firm research contacts exist with many other psychologists and psychiatrists, including prof. E.J. Wagenmakers, and dr. R. van Bork (Psychology UvA), prof. P. de Jonge and dr. L. Bringmann (Psychology RUG), and prof. K. Kendler (Psychiatry Richmond USA), member of the DSM-5 review committee.
- The applicant maintains working relationships with numerous statisticians and data scientists, including prof. P. Grünwald (CWI Amsterdam) and prof. R. Stolk (Epidemiology UMCG and head of the information technology centre CIT at RUG), as well as other data science experts in the research group of CIT. Moreover, the applicant has research ties with other philosophers who focus on data science, e.g., prof. G. Wheeler (Frankfurt) and prof. K. Genin (Tübingen).
- There are strong connections with prof. D. Borsboom (Psychology UvA), director of the Social and Behavioural Data Science Centre of the University of Amsterdam, and one of the initiators of the network approach in psychopathology.
- The applicant was responsible for the methodology section of the successful NWO Gravitation proposal “SCOOP” and is involved in maintaining its data infrastructure. On data infrastructures for social science, he is in regular contact with prof. C. Aarts (RUG) and others from the SCOOP consortium.

Implementation

Implementation consists in publications, workshops, and guidelines, which all take their cue from existing concerns over the justification of data science methods within psychopathology and beyond.

- The results of the project will be disseminated through joint publications in journals for psychopathology and statistical methodology. The applicant has ample experience with such collaborations. Joint research will be supported by embedding the PhD students and postdocs at the affiliated institutions.

- The project includes three focused workshops on the intersection of philosophy, data science, and psychology or psychiatry, to which relevant partners from social science contexts will be invited.
- The applicant will produce guidelines for researchers working with data science methods that will be advertised and distributed widely. These guidelines aim to make researchers aware of ways in which specific uses of data science methods import suppositions into their research, and it will offer concrete suggestions on how to report on that.

Societal relevance

Contacts and implementation

The applicant is actively involved in outreach activities and can make use of existing contacts to reach professionals and the general public.

- The results of the project will be disseminated to psychiatric professionals. Ties have already been forged with psychiatrists using data science in mental health care policy, like I-SHARED, and further opportunities will emerge from the engagement of clinical researchers.
- The applicant acts as advisor to Dutch courts, through regional and national training, course design, and professional standards. The use of data science in law is becoming increasingly important. Supported by the central training facility of the Dutch courts SSR, the applicant will offer expertise on data science. The use of data science methods in fraud detection and forensic investigations are a case in point, and so are the prediction methods used by professionals advising the police and judicial system, e.g., risk assessments for the purpose of public safety and reintegrating ex-detainees.
- An understanding of data science will help businesses to use machine learning more responsibly. To this aim the applicant has established contacts with housing market analyst Brainbay.
- A fourth workshop at the end of the project will focus on the communication of data science results, and will include a program part for media partners. To report and critically assess the role of data science in society, science journalists need a better understanding of how the methods work. Contact has been established with the secretary of the Dutch Association of Journalists (NVJ) and with journalists reporting on data science.
- A final outreach item is a set of three documentary films of 5-10 minutes for 14-16 year olds, about the possibilities and risks of data science in daily life, e.g., in smartphones, social media, and online shopping. through the online course “Wetenschapper in de Klas”, distributed by the Pre-University Academy of the RUG, which reached over 10,000 children aged 10-12. Together with the Academy and with experts on secondary school teaching, the applicant will produce the short films and the accompanying lecture materials. The Pre-University Academy will help distribute the films to schools nation-wide. A documentary film maker who previously lectured on new media and specializes in info-documentaries has already been found.

Literature

Aafjes-van Doorn, K. et al (2021). A scoping review ofmachine learning in psychotherapy research. Psychotherapy Research 31:1, pp. 92–116.

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

Angwin, J., S. Mattu, J. Larson and L. Kirchner (2016). Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks. ProPublica, retrieved August 2018.

Bacon, F. (1620/2014). Novum Organum. J. Devey (ed.), Gutenberg EBook #45988.

Barnett, V. (1999). Comparative Statistical Inference, Wiley Series in Probability and Statistics. Wiley.

Barocas, S. and Selbst, A.D. (2016). Big Data's Disparate Impact, California Law Review 671, 62 pages.

Beaulieu, A. (2004). From Brainbank to Database: The Informational Turn in the Study of the Brain. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 35(2), pp. 367–90.

Beijers, L., Wardenaar, K.J., Lamers, F., Romeijn, J.W., Loo, H.M. van, Schoevers, R.A. (2020). Investigating Data-driven Biological Subtypes of Psychiatric Disorders using Specification-Curve Analysis, Psychological Medicine, pp. 1-12.

Berger J. O. and Wolpert, R.L. (1988). The Likelihood Principle, 2nd edn. Institute of Mathematical Statistics, Hayward.

Birnbaum, A. (1962). On the foundations of statistical inference. Journal of the American Statistical Association 57(298), pp. 269–306.

Blaauw, F. (2017). The non-existent average individual. Dissertation Rijksuniversiteit Groningen.

Bork, R. van (2019). Interpreting psychopetric models. Dissertation University of Amsterdam. DOI: 10.31237/osf.io/x6a7s.

Borsboom, D. And A.O.J. Cramer (2013). Network Analysis: An Integrative Approach to the Structure of Psychopathology. Annual Review of Clinical Psychology 9(1), pp. 91–121.

Borsboom D., J.W. Romeijn and J.M. Wicherts (2008). Measurement invariance versus selection invariance: is fair selection possible? Psychological Methods 13(2), pp. 75–98.

Breiman L. (2001) "Statistical Modeling: The Two Cultures" *Statistical Science* 16(3), 199–231.

Bringmann, L. F. et al (2019). What do centrality measures measure in psychological networks? Journal of Abnormal Psychology. DOI:10.1037/abn0000446

Carnap, R. (1950). The Logical Foundations of Probability. University of Chicago Press.

Carnap, R. (1952). The Continuum of Inductive Methods. University of Chicago Press.

Cartwright N. and Hardie, J. (2012). Evidence-Based Policy: A Practical Guide to Doing It Better. Oxford University Press.

Caruana, R. (2017) "Friends Don't Let Friends Deploy Black-Box Models", FATML 2017 talk.

Chalupka, K., Eberhardt, F. and Perona P. (2017). Causal feature learning: an overview. Behaviormetrika 44:137–164.

Chekroud, A. M. et al (2016). Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psyciatry 3(3), pp. 243–50.

Ching T. et al (2018). Opportunities and obstacles for deep learning in biology and medicine. J. R. Soc. Interface 15: 20170387. DOI: 10.1098/rsif.2017.0387.

Claeskens, G. and Hjort, N. L. (2008). Model selection and model averaging. Cambridge University Press.

Cooper, R. (2014), Diagnosing the diagnostic and statistical manual of mental disorders. London: Karnac Books.

Corfield, D. (2010). Varieties of Justification in Machine Learning. Minds and Machines 20:291–301.

Creel, K. A. (2020). Transparency in Complex Computational Systems. Philosophy of Science 87, pp. 568–589.

Daston, L.J. and Galison, P. (2007). Objectivity. The MIT Press.

Dawid, A.P. and Stone, M. (1982). The functional-model basis of fiducial inference. Annals of Statistics 10, pp. 1054–1067.

De Finetti, B. (1937/1964). Foresight: its Logical Laws, Its Subjective Sources. Translation in Studies in Subjective Probability, eds. H. Kyburg and H. Smokler, Wiley.

Diggle, P.J. (2015). Statistics: a data science for the 21st century. Journal of the Royal Statistical Society A, 178:4, pp. 1–18.

Doshi-Velez, F., Kortz, M. (2017). Accountability of AI Under the Law: The Role of Explanation, ArXiv arXiv:1711.01134v2 [cs.AI].

Douglas, H. (2009). Science, Policy, and the Value-Free Ideal. Pittsburgh University Press.

Efron, B. and Morris, C. (1973). Stein's Estimation Rule and Its Competitors – An Empirical Bayes Approach. Journal of the American Statistical Association 68(341), pp. 117–130.

Elsayed, G. et al. (2018). Adversarial Examples that Fool both Human and Computer Vision. arXiv:1802.08195v2 [cs.LG] 27 Feb 2018.

Eronen, M. and J.W. Romeijn (2020), Philosophy of Science and the Formalization of Psychological Theory, Theory & Psychology, Vol. 30(6), pp. 786 –799.

Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.

ACM FAccT (2021). https://facctconference.org/network/, retrieved September 2021.

Fisher, R.A. (1956). Statistical Methods and Scientific Inference, New York: Hafner, 3rd edition.

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

Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry* 31(4), pp. 271-288.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A. and Rubin, D. (2013). Bayesian Data Analysis, revised edition. Chapman & Hall/CRC.

George, S. V., Y. K. Kunkels, S. Booij, and M. Wichers (2021). Uncovering complexity details in actigraphy patterns to differentiate the depressed from the non-depressed. Scientific Reports 11(1), DOI:10.1038/s41598-021-92890-w.

Goodfellow, I, Y. Bengio and A. Courville (2016). "Deep Learning". The MIT press.

Goodman, B. and Flaxman, S. (2017). "EU Regulations on Algorithmic Decision-Making and a 'Right to Explanation.'" AI Magazine, Fall 2017.

de Heide, R. and Grünwald, P.D. (2018). Why optional stopping is a problem for Bayesians. ArXiv:1708.08278v3.

Harman, G. and Kulkarni, S. (2007). Reliable Reasoning: Induction and Statistical Learning Theory. MIT press.

Hartmann, S., Gabbay, D. and Woods, J. (2011). Handbook of the History of Logic, Vol. 10: Inductive Logic. Elsevier: North Holland.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). The elements of statistical learning; data mining, inference, and prediction. Springer-Verlag.

Hey, T., S. Tansley and K. Tolle (2012). The Fourth Paradigm: Data-Intensive Scientific Discovery. E-Science and Information Management. Communications in Computer and Information Science. 317: 1.

Huttegger, S. (2017). The Probabilistic Foundations of Rational Leaning. Cambridge University Press.

Hutter, M. (2007). Universal Algorithmic Intelligence: A Mathematical Top→Down Approach. In Artificial General Intelligence, Springer, pp. 227-290.

Howson, C. (2000). The Problem of Induction. Oxford University Press.

Ioannides, J.P.A. (2005). Why Most Published Research Findings Are False. PloS-Medicine. DOI: 10.1371/journal.pmed.0020124.

Juengst, E., McGowan, M.L., Fishman, J.R. and Settersten, R. A. (2016). From "Personalized" to "Precision" Medicine: The Ethical and Social Implications of Rhetorical Reform in Genomic Medicine. *Hastings Center Report* 46, 21–33.

Kan, K. et al (20XX). The co-creation of a decision-aid for patients with depression: Combining data-driven prediction with patients' and clinicians' needs and perspectives. Manuscript.

Kelly, K. (1996). The Logic of Reliable Inquiry. Oxford University Press.

Kendler, K.S. (2012). The dappled nature of causes of psychiatric illness: replacing the organic–functional/hardware–software dichotomy with empirically based pluralism. Molecular Psychiatry 17, pp.377–388.

Kendler, K., J. Parnas and P. Zachar (2020). Levels of Analysis in Psychopathology. Cambridge University Press.

Kessler, R. C., van Loo, H., et al (2016). Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. Molecular Psychiatry 21(10), 1366-1371. DOI: 10.1038/mp.2015.198.

Kleinberg, J., S. Mullainathan and M. Raghavan (2016) "Inherent Trade-Offs in the Fair Determination of Risk Scores", arXiv:1609.05807 [cs.LG].

Kieseppä, I. A. (1997). Akaike Information Criterion, Curve-Fitting, and the Philosophical Problem of Simplicity. British Journal for the Philosophy of Science 48(1), pp. 21–48.

Kieseppä, I.A. (2001). Statistical Model Selection Criteria and the Philosophical Problem of Underdetermination. British Journal for the Philosophy of Science 52(4), pp. 761–794.

Kincaid, H. (2008). Do We Need Theory to Study Disease? Lessons from cancer research and their implications for mental illness. Perspectives in Biology and Medicine, 51(3), 367-378.

Kitcher, P. (2003). Science, Truth, and Democacy. Oxford University Press.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. Big Data & Society, April–June 2014: 1–12.

Kotov, R. et al (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A Dimensional Alternative to Traditional Nosologies. Journal of Abnormal Psychology 126:4, pp. 454–477.

Kuffner, T. and G. Young (2018). Philosophy of Science, principled statistical inference, and data science. Manuscript

Kuhn, T. (1962). The Structure of Scientific Revolutions. Chicago University Press.

Laan, M.J. van der, and Rose, S. (2011). Targeted Learning: Causal Inference for Observational and Experimental Data. Springer.

Latour, B. (1987). Science in action. Harvard University Press.

LeCun, Y. (2019). The epistemology of deep learning. Presentation at workshop Deep Learning: Alchemy or Science?, Institute for Advanced Study.

Leonelli, S. (2016) Data-Centric Biology: A Philosophical Study. Chicago University Press.

Leonelli, S. (2016). Locating Ethics in Data Science: Responsibility and Accountability in Global and Distributed Knowledge Production Systems. Phil. Trans. R. Soc. A 374 (2083): 20160122.

Leonelli, S. (2020). Learning from Data Journeys. In Leonelli, S. and N. Tempini, Data Journeys in the Sciences, Springer, pp. 1–24.

Lipton, Z. C. (2016). The Mythos of Model Interpretability. In Proceedings of the 2016 ICML Workshop on Human Interpretability in Machine Learning. New York, NY.

Longino, H. (1990). Science as Social Knowledge. Princeton University Press.

van Loo, H. and J.W. Romeijn (2015). Psychiatric comorbidity: fact or artifact? Theoretical medicine and bioethics 36(1), 41–60.

van Loo, H. M., & Romeijn, J. W. (2018). Measuring and defining: The double role of the DSM criteria for psychiatric disorders. *Psychological Medicine* 48(5), pp. 872-873.

van Loo, H. M., K. S. Kendler, J.W. Romeijn (2019). Changing the definition of the kilogram: insights for psychiatric disease classification. Philosophy, Psychiatry and Psychology 26(4), pp. 97–108.

Mackenzie, A. 2017. *Machine Learners: Archaeology of a Data Practice*. MIT Press.

Matthews, L.J., Turkheimer, E. (2021). Across the great divide: pluralism and the hunt for missing heritability. Synthese 198, 2297–2311.

Mayer-Schönberger, V. and K. Cukier, “Big Data: A Revolution That Transforms How we Work, Live, and Think”, Houghton Mifflin Harcourt.

Mooij, J. (2020). Joint Causal Inference from Multiple Contexts. Journal of Machine Learning Research 21, pp. 1–108.

Moretti, F. (2013). Distant Reading. London and New York: Verso.

Morey, R., J.W. Romeijn and J.N. Rouder (2016). The philosophy of Bayes’ factors. Journal of Mathematical Psychology 72, pp. 6–18.

Nevin L., on behalf of the *PLOS Medicine* Editors (2018). Advancing the beneficial use of machine learning in health care and medicine: Toward a community understanding. PLoS Med 15 (11): e1002708.

Neyman, J. and Pearson, E. (1967). Joint Statistical Papers. Cambridge University Press

Nguyen A., J. Yosinski, J. Clune (2015) “Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images”, CVPR '15, IEEE.

O’Neil, C. (2016). Weapons of Math Destruction. Penguin Books.

Open Science (2021). https://osf.io/. Retrieved September 2021.

Ortner, R. and H. Leitgeb (2011), “Mechanizing Induction”, in Handbook of the History of Logic: Inductive Logic, Ed. D. Gabbay, S. Hartmann and J. Woods, Elsevier, 719–772.

Paris, J.B. (1994). The uncertain reasoner’s companion. Cambridge University Press.

Paris, J. and Vencovská, A. (2015). Pure Inductive Logic. Cambridge University Press.

Patel, A.B., Nguyen, T. and Baraniuk, R.G. (2015). A Probabilistic Theory of Deep Learning. ArXiv: 1504.00641v [stat.ML].

PLoS Medicine Editors (2018). Advancing the beneficial use of machine learning in health care and medicine: Toward a community understanding. PLoS Medicine 15(11): e1002708.

Poland, J. and Tekin, S. (2017). Introduction: Psychiatric Research and Extraordinary Science. In Extraordinary science and psychiatry. The MIT Press.

Popper, K. (1934/1959). The Logic of Scientific Discovery. Routledge.

Putnam, H. (1963) " 'Degree of confirmation' and inductive logic", in The Philosophy of Rudolf Carnap, ed. P. Schilpp, Open Court, La Salle, 761–783.

Rahimi, A. and B. Recht (2017). Reflections on random kitchen sinks. Acceptance speech for Test of Time Award at the 30th International Conference on Neural Information Processing Systems.

Romeijn, J.W. (2004). Hypotheses and Inductive Predictions. Synthese 141(3), pp. 333–364

Romeijn, J.W. (2005). Bayesian Inductive Logic. Dissertation University of Groningen.

Romeijn, J.W. (2011). Statistics as inductive logic. In Handbook for the Philosophy of Science: Philosophy of Statistics, eds. P. Bandyopadyay and M. Forster, Elsevier, 751–774.

Romeijn, J.W. (2014a). Philosophy of Statistics", Stanford Encyclopedia, ed. E.N. Zalta, pp. 1–86.

Romeijn, J.W. (2014b). Humanities' New Methods: a reconnaissance mission. In The Making of the Humanities III, ed. R. Bod, J. Maat and Th. Weststeijn. Amsterdam: Amsterdam University Press, pp. 527–539.

Romeijn, J.W. (2017). Implicit complexity. Philosophy of Science 84(5), pp. 797-809.

Romeijn, J.W. (2020). Psychiatric classification: an a-reductionist perspective. In Conceptual Issues in Psychiatry, ed. K. Kendler and J. Parnas, Cambridge University Press.

Ross, L. (20XX). Explanation in Contexts of Causal Complexity: Lessons from Psychiatric Genetics. Manuscript.

Roscher, R. et al (2020). Explainable Machine Learning for Scientific Insights and Discoveries. IEEE Access. DOI: 10.1109/ACCESS.2020.2976199.

Savage, L.J. (1954). The Foundations of Statistics. Dover books.

Schaffner, K. F. & Tabb, K. (2015). Varieties of social constructionism and the problem of progress in psychiatry. In K. S. Kendler & J. Parnas (eds.), International perspectives in philosophy and psychiatry. Philosophical issues in psychiatry III: The nature and sources of historical change. Oxford University Press, pp. 85–106.

Schölkopf, B. and Smola, A.J. (2001). Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. The MIT press.

Schurz, G. (2019). Hume's Problem Solved: The Optimality of Meta-Induction. The MIT press.

Seidenfeld, T. (1979). Philosophical Problems of Statistical Inference: Learning from R.A. Fisher. Dordrecht: Reidel.

Seidenfeld, T. (1992). R.A. Fisher's Fiducial Argument and Bayes Theorem. Statistical Science 7(3), pp. 358–368.

Shalev-Shwartz, S. (2019). Surrogates. Presentation at workshop Deep Learning: Alchemy or Science?, Institute for Advanced Study

Skyrms, B. (1996). Carnapian Inductive Logic and Bayesian Statistics. In Statistics, Probability and Game Theory: Papers in Honor of David Blackwell, pp. 321–336.

Solomonoff, R. (1964). A Formal Theory of Inductive Inference, Part I. Information and Control 7:1, pp. 1–22.