Article

Approximate and Situated Causality in Deep Learning

Jordi Vallverdú

Philosophy Department, Universitat Autònoma de Barcelona, 08193 Bellaterra (BCN), Catalonia, Spain;
jordi.vallverdu@uab.cat; Tel.: +34935811618

Abstract: Causality is the most important topic in the history of western science, and since the beginning of the statistical paradigm, its meaning has been reconceptualized many times. Causality entered into the realm of multi-causal and statistical scenarios some centuries ago. Despite widespread critics, today deep learning and machine learning advances are not weakening causality but are creating a new way of finding correlations between indirect factors. This process makes it possible for us to talk about approximate causality, as well as about a situated causality.

Keywords: causality; deep learning; machine learning; counterfactual; explainable AI; blended cognition; mechanisms; system

1. Causalities in the 21st Century

In classical western philosophy, causality was considered an obvious observation of the divine regularities which were ruling Nature. From a dyadic truth perspective, some events were true while others were false, and those which were true strictly followed Heaven's will. That ontological perspective allowed early Greek philosophers (inspired by Mesopotamian, Egyptian, and Indian scientists) to define causal models of reality with causal relations deciphered from a single origin, the arche (ἀρχή). Anaximander, Anaximenes, Thales, Plato, or Aristotle, among others, created different models about causality, all of them connected by the same idea: hazard or nothingness was not possible. Despite the fact that those ideas were defended by atomists (who considered Nature with both hazard and void), any trace of them was deleted from the research. On the other hand, eastern philosophers departed from the opposite ontological point of view: At the beginning there was nothingness, and the only true reality is the continuous change of things [1]. For Buddhist (using a four-valued logic), Hindu, Confucian, or Taoist philosophers, causality was a reconstruction of the human mind, which is also a non-permanent entity. Therefore, the notion of causality is ontologically determined by situated perspectives about information values [2], which allowed and fed different and fruitful heuristic approaches to reality [3,4]. Such situated contexts of thinking shape the ways by which people perform epistemic and cognitive tasks.

These ontological variations can be justified and fully understood once we assume the Duhem–Quine thesis, that is, that it is impossible to test a scientific hypothesis in isolation, because an empirical test of the hypothesis requires one or more background assumptions (also called auxiliary assumptions or auxiliary hypotheses). Therefore, the history of the idea of causality changes coherently across the geographies and historical periods, entering during the late 19th century into the realm of statistics and, later in the 20th century, in multi-causal perspectives [5]. The statistical nature of contemporary causality has been involved into debates between schools, mainly Bayesians and a broad range of frequentist variations. At the same time, the epistemic thresholds have been changing, as the recent debate about statistical significance has shown, desacralizing the *p*-value. The most recent and detailed academic debate on statistical significance was extremely detailed into the

#1 Supplement of Volume 73, 209 of the journal *The American Statistician*, released in March 20th 2019. But during the last decades of the 20th century and the beginning of the 21st century, computational tools have become the backbone of cutting-edge scientific research. After the great advances produced by machine learning (ML) techniques, several authors have asked themselves whether ML can contribute to the creation of causal knowledge. We will provide an answer to this question in the next section.

2. Deep Learning, Counterfactuals, and Causality

It is in this context, where the statistical analysis rules the study of causal relationships, that we find the challenge to machine learning and deep learning that considers them as not suitable tools for the advance of causal and scientific knowledge. The most known and debated arguments come from the eminent statistician Judea Pearl [6,7], and have been widely accepted. The main idea is that machine learning does not create causal knowledge because it lacks the skill of managing counterfactuals, and following his exact words, [6] page 7: "Our general conclusion is that human-level AI cannot emerge solely from model-blind learning machines; it requires the symbiotic collaboration of data and models. Data science is only as much of a science as it facilitates the interpretation of data—a two-body problem, connecting data to reality. Data alone are hardly a science, regardless how big they get and how skillfully they are manipulated". What he is describing is the well-known problem of the black-box model: we use machines that process very complex amounts of data and provide some extractions at the end. As it has been called, it is a GIGO (Garbage In, Garbage Out). Following this line of argument, it could be affirmed that GIGO problems are computational versions of the Chinese room thought experiment: the machine can find patterns but without real and detailed causal meaning. This is what Pearl criticizes: the blind use of data for establishing statistical correlations instead of describing causal mechanisms. We will analyze in the next sections the problematic tensions between correlations and causal patterns in sets of data using deep learning methods.

2.1. Deep Learning Is Not A Data-Driven but A Context-Driven Technology: Made by Humans for Humans

Most of the epistemic criticisms against AI are always repeating the same idea: machines are still not able to operate as humans do. The idea is always the same: computers are operating with data using a blind semantic perspective that makes it not possible that they understand the causal connections between data. It is the definition of a black-box model. But here is where we find the first problem: deep learning (DL) is not the result of automated machines creating, by themselves, search algorithms and after it, evaluating them as well as their results. DL is designed by humans, who select the data, evaluate the results, and decide the next step into the chain of possible actions. At the epistemic level, the decision about how to interpret the validity of DL results is under human evaluation [8]. Even the latest trends in AGI design include causal thinking, as the DeepMind team has recently detailed [9], and with explainable properties. The exponential growth of data and their correlations has been affecting several fields, especially epidemiology. Initially, it can be expressed by the agents of some scientific community as a great challenge, in the same way that astronomical statistics modified the Aristotelian–Newtonian idea of a physical cause, but with time, the research field accepts new ways of thinking. Consider also the revolution of computer proofs in mathematics and the debates that these techniques generated among experts.

In that sense, DL is just providing a situated approximation to reality using correlational coherence parameters designed by the communities that use them. It is beyond the nature of any kind of machine learning to solve problems only related to human epistemic envisioning: let us take the long, unfinished, and even disgusting debates among the experts of different statistical schools [5]. This is true because data do not provide or determine epistemology, in the same sense that groups of data do not provide the syntax and semantics of the possible organization systems to which they can be assigned. Any connection between the complex dimensions of any event expresses a possible epistemic approach, which is a (necessary) working simplification. We cannot understand the world using the world itself, in the same way that the best map is not a 1:1 scale map, as Borges wrote (1946,

On Exactitude in Science): "…In that Empire, the Art of Cartography attained such Perfection that the map of a single Province occupied the entirety of a City, and the map of the Empire, the entirety of a Province. In time, those Unconscionable Maps no longer satisfied, and the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it. The following Generations, who were not so fond of the Study of Cartography as their Forebears had been, saw that that vast Map was Useless, and not without some Pitilessness was it, that they delivered it up to the Inclemencies of Sun and Winters. In the Deserts of the West, still today, there are Tattered Ruins of that Map, inhabited by Animals and Beggars; in all the Land there is no other Relic of the Disciplines of Geography".

Then, DL cannot follow a different information processing process, a specific one completely different from those run by humans. As with any other epistemic activity, DL must include different levels of uncertainties if we want to use it [10]. Uncertainty is a reality for any cognitive system, and consequently, DL must be prepared to deal with it. Computer vision is a clear example of that set of problems [11]. Kendall and Gal have even coined new concepts to allow the introduction of uncertainty into DL: homoscedastic, and heteroscedastic uncertainties (both aleatoric) [12]. The way used to integrate such uncertainties can determine the epistemic model (which is a real cognitive algorithmic extension of ourselves). For example, the Bayesian approach provides an efficient way to avoid overfitting, allows the ability to work with multi-modal data, and makes possible their use in real-time scenarios (as compared to Monte Carlo approaches) [13]; or even better, some authors are envisioning Bayesian deep learning [14]. Dimensionality is a related question that also has a computational solution, as Yosuhua Bengio has been exploring during the last decades [15–17].

In any case, we cannot escape from the informational formal paradoxes, which were well-known at a logical and mathematical level once Gödel explained them; they just emerge in this computational scenario, showing that artificial learnability can also be undecidable [18]. Machine learning is dealing with a rich set of statistical problems, those that even at a biological level are calculated at approximate levels [19], a heuristic that is also being implemented into machines. Such an open range of possibilities, as well as the existence of mechanisms like informational selection procedures (induction, deduction, and abduction), makes it possible to use DL in a controlled but creative operational level.

New Reasoning and DL

As a surprising fact, DL, which was initially strongly related to inductive techniques, is allowing the automation of still unexplored ways of thinking, like abduction [20,22].Abduction and induction are both strictly related forms of defeasible reasoning from effects to causes. This already happened some decades before with Bayesianism, which experienced a boost in its application thanks to the new availability of computational power. Currently, abductive reasoning can be the key to data management in the era of DL. Thanks to the raw computational power and the use of fine statistical methods, computational abduction can provide a delicate way of finding new causal relationships between sets of data [23]. The work of Vladimir Vapnik is enlightening such possibilities, as in [24], where he introduces a statistical approach to statistical inference that in philosophy is also called "the Duck Test" (understood as is described by its usual expression: "if it looks like a duck, swims like a duck, and quacks like a duck, then it probably is a duck"). In that work, Formula (13), Vapnik considers such new kind of inference based on statistical invariants (see Formula (13)). Since it is valid for any predicate (any function $\psi(x)\in L_2$) one can construct as many statistical invariants (defining properties of class) as one wants. The increase in the rate of convergence is not the only benefit of such an approach, because it also allows holistic descriptions using privileged information [25]. This use of privileged information, following the formalization of abduction, is close to the way by which humans' reason, combining strategies: holistic descriptions, use of complex models, etc. Finally, in relation to these ideas, such an approach also allows Vapnik to create non-inductive inferences, replacing it by what he calls *transductive inference*, much more predictive than classic generative learning models [26].

On the other hand, there is a new, or somehow old, taking into consideration the precedents, way of achieving new machine learning: a combination of neural-based models and symbolic programs. The recent Neuro-Symbolic Concept Learner (NS-CL), made by MIT, IBM, and Deepmind experts [27] is a reliable way of checking such a possibility.

2.2. Deep Learning Is Already Running Counterfactual Approaches

The second big Pearl criticism against DL focuses on its incapacity of integrating counterfactual heuristics. First, we must affirm that counterfactuals do not warrant with precision any epistemic model, just add some value (or not). From a classic epistemic point of view, counterfactuals do not provide a more robust scientific knowledge: a quick look at the last two thousand years of both western and eastern sciences can give support to this view [4]. Even going beyond, I affirm that counterfactuals can block thinking once it is structurally related to a close domain or paradigm of well-established rules; otherwise, it is just fiction or an empty mental experiment. Counterfactuals are a fundamental aspect of human reasoning, and their algorithmic integration is a good idea [28]. But at the same time, due to the problem of underdetermination, counterfactual thinking can express completely wrong ideas about reality. DL cannot have an objective ontology that allows it to design a perfect epistemological tool: because of the huge complexity of the involved data as well as for the necessary situatedness of any cognitive system. Uncertainty would not form part of such counterfactual operationability [29], as it should be ascribed to any not-well-known but domesticable aspect of reality; nonetheless, some new ideas do not fit with the whole set of known facts, the current paradigm, nor the set of new ones. This would position us into a sterile no man's land, or even block any sound epistemic movement. But humans are able to deal with it, even going beyond [30]. Opportunistic blending and creative innovation are part of our set of most valuable cognitive skills [31].

2.3. DL Is Not Magic Algorithmic Thinking (MAT)

Our third and last analysis of DL characteristics is related to its explainability. Despite the evidence that the causal debate is beyond any possible resolution provided by DL, because it belongs to ontological perspectives that require a different holistic analysis, it is clear that the results provided by DL must be not only coherent but also explainable, otherwise we would be in front of a new algorithmic form of magic thinking. By the same reasons that DL cannot be just mere a complex way of curve fitting, it cannot become a fuzzy domain beyond human understanding. Some attempts are being held to prevent us from this, most of them are rules by DARPA: Big Mechanisms [32] or eXplainable Artificial Intelligence (XAI) [33], as Figure 1 details. An image from the DARPA website:
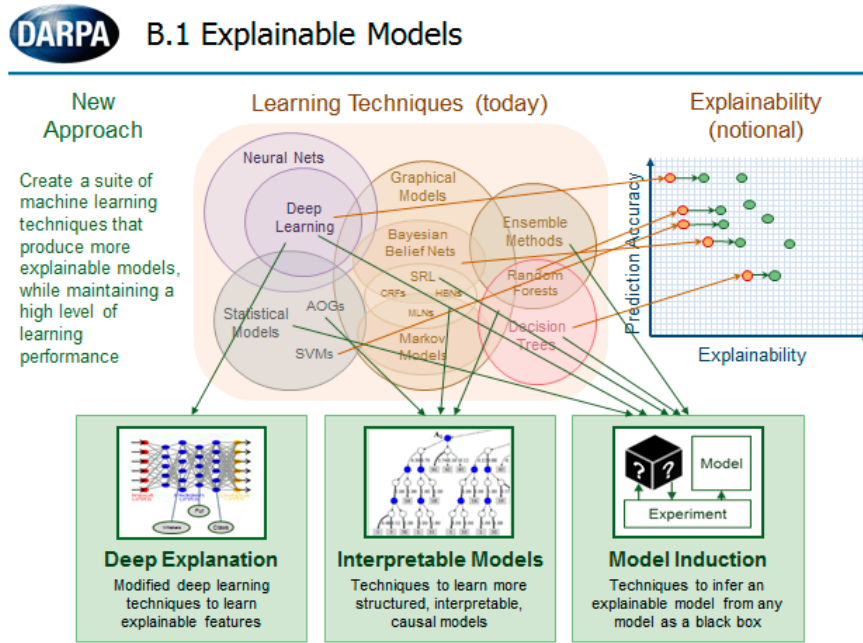
Figure 1. Explainability in deep learning (DL).

Again, the approaches detailed in Figure 2 answer to a request: how can we adapt new epistemic tools to our cognitive performative thresholds and characteristics. There is not a bigger revolution from a conceptual perspective than the ones that happened during the Renaissance with the use of telescopes or microscopes. DL systems are not running by themselves interacting with the world, automatically selecting the informational events to be studied, or evaluating them in relation to a whole universal paradigm of semantic values.
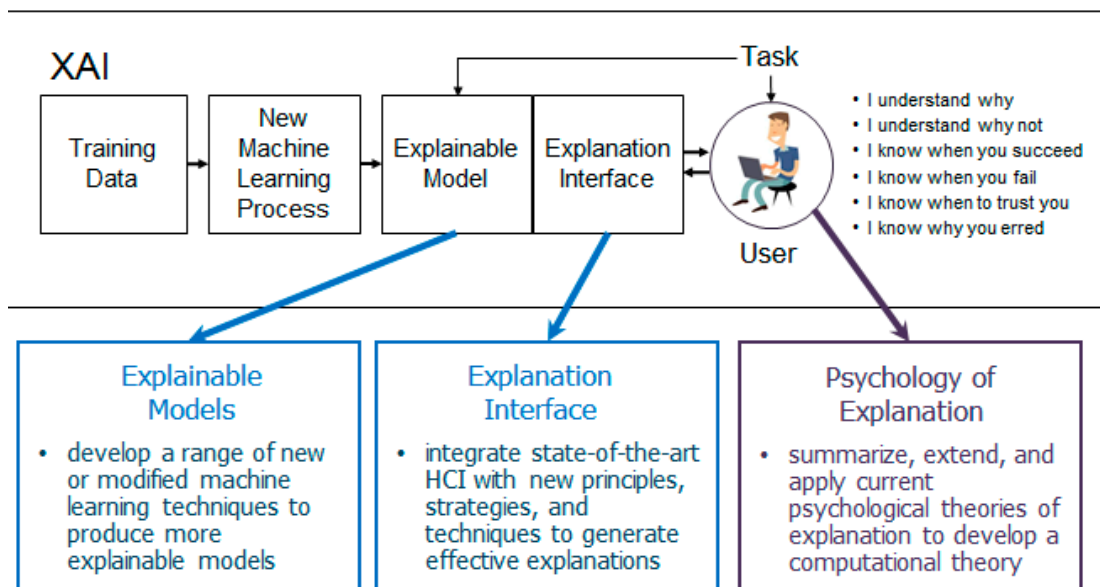

Figure 2. EXplainable Artificial Intelligence (XAI) from DARPA.

Singularity debates are useful for exploring possible conceptual frameworks and must be held, but at the same time are at risk of becoming fallacious fatalist arguments against current knowledge. Today, DL is a tool used by experts in order to map new connections between sets of data. Epistemology is not an automated process, despite the minor and naïve attempts to achieve it. Knowledge is a complex set of explanations related to different systems that are integrated dynamically by networks of epistemic (human, still) agents who are working with AI tools. Machines could postulate their own models, true, but the mechanisms to verify or refine them would not be beyond any mechanism different from the used previously by humans: data do not express by themselves some pure nature, but offers different system properties that need to be classified in order to obtain knowledge. This obtaining is somehow a creation based on the epistemic and body situatedness of the system.

2.4. DL Affects Scientific Thinking

Without reliable causal connections between events, the whole scientific project trembles. The data deluge of new e-Science contexts has emerged as a menace for current rational thinking [34]. However, the statistical nature of DL has made it more necessary to consider how spurious correlations could infer the use of big data [35]. If, during the 19th century, statistical thinking revolutionized the causal paradigm, the new set of statistical tools and the informational context of the implementation of DL has faced contemporary science in front of a new revision process. The old worries about the lack of causal precision are again between us, but also the intuition about the impossibility of checking the internal mechanisms by which such statistical connections are obtained is today a serious worry. Well, even the opposite attitude is worrying: just consider The COURSERA course on Deep Learning by the influential Andrew Ng, in which he affirmed "If you don't understand the calculus, don't worry!", and became automatically a worldwide famous meme. This is the current spirit among DL practitioners, a consequentialist one: forget truth, we want only to apply it and look for new results than can feed our pockets/CV. However, there are other successful approaches, like those of Léon Bottou, from Facebook AI research: *invariance* across environments is related to causation. Here, we face the supradetermination problem (or Duhem–Quine hypothesis), because even invariance across environments could be the result of a wholly wrong paradigm design, as happened several times all throughout the history of western science. In any case, the deep changes are not the statistical black boxes, but our pragmatic condescendence with (un)reliable knowledge: is it enough just in case we feel comfortable at the epistemic level? As we can infer from what happened with Dijkstra and his GOTO programming debate, pragmatists always win. Perhaps this is the true answer to the epistemic debates about DL and causality: do not overestimate causality, because perhaps we have never truly achieved it.

2.5. Recent Attempts to Obtain Causal Patterns in DL

In this last sub-section I want to explore an insightful and recent approach based on algorithmic calculus [36,37], rooted in Somolonoff's inductive inference ideas. What these authors have tried is to teach machine learning cause and effect, using a combination of algorithmic perspective that supersedes mere operatively successful but "blind" statistical approaches. Using symbolic computation together with counterfactuals, perturbation analysis, as well as with the combinatorial power of current approaches to DL, they work towards the achievement of causal understanding in machine learning. In this approach, the parsing of the event or object under study is analyzed by a group of programs. Such algorithmic probability deconvolves the interacting mechanisms looking for a final macro-reconstruction. This is the most promising current approach to causal understanding at the computational level, although some pending issues must be considered: in real scenarios, causal processes are related to sets of objects that generate multidimensional responses in relation to the local conditions: think for example on gene expression according to the environmental variations [38]. This dynamic nature of causality implies the possibility of defining an object's properties according to local variations, that is, in relation to the affordability of such originators for the generation of different patterns. At the same time, at an ontological level, causal understanding

responds to different sets of meta-evaluations about Nature's reality. A fundamental mechanistic approach must also consider the system in which we analyze such a causal pattern. As we will see in Section 4, the delimitation of analysis thresholds modifies the notion of causality, something that completely changes the mental horizons of statisticians doing epidemiological analysis. My remark is epistemological: we need to define the smoothness between causal relations of sets of data (the level of required causal connectivity), as well as a method for evaluating/selecting the causal weight and length of multicausal events. At this level of analysis, DL needs not also a better algorithmic approach, like that cleverly suggested by the cited authors, but also an epistemic algorithmic approach, which introduces and combine variables about the necessary economy of causes, truth confidence, risk evaluations about generated knowledge, and global coherence (with the global set of epistemic knowledge, the scientific paradigm). For some obvious reasons, these processes must be supervised by humans at specific checking points of the whole process, or only at the end, in case of looking for more extreme (and possibly mistaken) results.

3. Extending Bad and/or Good Human Cognitive Skills Through DL

It is beyond any doubt that DL is contributing to improving the knowledge in several areas, some of them very difficult to interpret because of the nature of obtained data, like neuroscience [39]. These advances are expanding the frontiers of verifiable knowledge beyond classic human standards. However, even in that sense, they are still explainable. Anyhow, humans are the ones who fed DL systems with scientific goals, provide data (from which to learn patterns), and define quantitative metrics (in order to know how close you are to success). At the same time, are we sure that it is not our biased way to deal with cognitive processes that is the mechanism that allows us to be creative? For this reason, some attempts to reintroduce human biased reasoning into machine learning are being explored [40]. This re-biasing [41] even reapplies emotional-like reasoning mechanisms [42,43].

My suggestion is that after the great achievements following classic formal algorithmic approaches, it now time for DL practitioners to expand the horizons, looking into the great power of cognitive biases.

For example, machine learning models with human cognitive biases are already capable of learning from small and biased datasets [44]. This process reminds me of the role of the Student test in relation to frequentist ideas, always requesting large sets of data until the creation of the *t*-test, something that could be applied now in the context of machine learning.

In [44], the authors developed a method to reduce the inferential gap between human beings and machines by utilizing cognitive biases. They implemented a human cognitive model into machine learning algorithms and compared their performance with the currently most popular methods, naïve Bayes, support vector machine, neural networks, logistic regression, and random forests. This even could make possible one-shot learning systems [45]. Approximate computing can boost the potentiality of DL, diminishing the computational power of the systems as well as adding new heuristic approaches to information analysis.

Finally, a completely different type of problem, but also important, is how to reduce the biased datasets or heuristics we provide to our DL systems [46] as well as how to control the biases that make us not able to interpret DL results properly [47]. Obviously, if there is any malicious value related to such bias, it must be also controlled.

4. Causality in DL: The Epidemiological Case Study

Several attempts have been implemented in order to allow causal models in DL, like [48] and the structural causal model (SCM) (as an abstraction over a specific aspect of the CNN, a Convolutional Neural Network). We also formulate a method to quantitatively rank the filters of a convolution layer according to their counterfactual importance), or the temporal causal discovery framework (TCDF, a deep learning framework that learns a causal graph structure by discovering causal relationships in observational time series data) by [49]. My attempt here will be twofold: (1) first, to consider the value of "causal data" for epistemic decisions in epidemiology; and (2) second, to look at how DL could fit, or not, with those causal claims in the epidemiological field.

4.1. Does Causality Affect Epidemiological Debates At All?

According to the field reference [50], MacMahon and Pugh created one of the most frequently used definitions of epidemiology: "Epidemiology is the study of the distribution and determinants of disease frequency in man". Note the absence of the term 'causality' and, instead, the use of the one of 'determinant'. This is the result of the classic prejudices of Hill in his paper of 1965: "I have no wish, nor the skill, to embark upon philosophical discussion of the meaning of 'causation'. The 'cause' of illness may be immediate and direct; it may be remote and indirect underlying the observed association. But with the aims of occupational, and almost synonymous preventive, medicine in mind the decisive question is where the frequency of the undesirable event B will be influenced by a change in the environmental feature A. How such a change exerts that influence may call for a great deal of research, However, before deducing 'causation' and taking action we shall not invariably have to sit around awaiting the results of the research. The whole chain may have to be unraveled or a few links may suffice. It will depend upon circumstances." After this philosophical epistemic positioning, Hill numbered his nine general qualitative association factors, also commonly called "Hill's criteria" or even, which is frankly sardonic, "Hill's Criteria of Causation". For such epistemic reluctances, epidemiologists abandoned the term "causation" and embraced other terms like "determinant", "determining conditions", or "active agents of change". For that reason, recent research has claimed for a pluralistic approach to such complex analysis [51]. As a consequence, we can see that even in a very narrow specialized field like epidemiology the meaning of 'cause' is somehow fuzzy. Once medical evidence showed that causality was not always a mono-causality [52,53] but, instead, the result of the sum of several causes/factors/determinants, the necessity of clarifying multi-causality emerged as a first-line epistemic problem. It was explained as a "web of causation" [54]. Some debates about the logic of causation and some Popperian interpretations were held during sevreal decades [55]. Pearl himself provided a graphic way to adapt human cognitive visual skills to such new epidemiological multi-causal reasoning [56], as well as do-calculus [57], and directed acyclic graphs (DAGs) are also becoming a fundamental tool [58,59]. DAGs are commonly related to randomized controlled trials (RCT) for assessing causality. RCT are not a gold standard beyond any critic, because as [60] affirmed, RCT are often flawed, mostly useless, although clearly indispensable (it is not so uncommon that the same author claims against classic *p*-values suggesting a new 0,005, [61]). Krauss has even defended the impossibility of using RCT without biases [62], although some authors defend that DAGs can reduce RCT biases [63].

There is, however, a real case that can show us a good example of the weight of causality in real scientific debates. Consider the debates about the relation between smoking and lung cancer. As recent as 1950 the causal connections between smoking and lung cancer were explained [64]. Far from being accepted, however, these results were contradicted by the tobacco industry using scientific experimental regression. Perhaps the most famous generator of silly counterarguments was R.A. Fisher, the most important frequentist researcher of the 20th century. In 1958 he published a paper in *Nature,* in which he affirmed that all connections between tobacco smoking and lung cancer were due to a false correlation. Even more: With the same data it could be inferred that "smoking cigarettes was a cause of considerable prophylactic value in preventing the disease, for the practice of inhaling is rare among patients with cancer of the lung that with others" (p. 596). Two years later he was saying similar silly things in a highly-rated academic journal. He even affirmed that Hill tried to plant fear into good citizens using propaganda, and entering misleadingly into the thread of overconfidence. The point is: did have Fisher real epistemic reasons for not accepting the huge amount of existing causal evidence against tobacco smoking? No, and we are not affirming the consequent after collecting more data not available during Fisher's life. He had strong causal evidence but he did not want to accept them. Still today, there is evidence that shows how causal connections are field biased, again with tobacco or the new e-cigarettes. For such reasons, it is not clear what the real value of counterfactuals is, as required by R.A. Fischer himself, as a way to evaluate a set of data under a hypothesis. Even the same counterfactual description is related to the conceptual paradigm in which such statements accomplish the role and scope of acts as counterfactuals, being consequently related to some ontological perspective that offers a stable and accepted validity. In this sense, the

use of counterfactuals is part of a local and scalable epistemology. As a section conclusion, it can be affirmed that causality has strong specialized meanings and can be studied under a broad range of conceptual tools. The real example of tobacco controversies offers such long temporal examples.

4.2. Can DL Be of Some Utility for the Epidemiological Debates on Causality?

The second part of my argumentation will try to elucidate whether DL can be useful for the resolution of debates about causality in epidemiological controversies. The answer is easy and clear: yes. However, it is directly related to a specific idea of causality as well as of a demonstration. For example, a machine learning approach can be found to enable evidence-based oncology. Thus, digital epidemiology is a robust update of previous epidemiological studies. The new possibilities of finding new causal patterns using bigger sets of data is surely the best advantages of using DL for epidemiological purposes [65]. Besides, such data are the result of integrating multimodal sources, like visual sources combined with classic informational sources [66], but the future, with different modes and more data capture devices could integrate smell, taste, movements of agents, etc. Deep convolutional neural networks can help us, for example, to estimate environmental exposures using images and other complementary data sources such as cell phone mobility and social media information. Combining fields such as computer vision and natural language processing, DL can provide a way to explore new interactions still opaque to us.

Despite the possible benefits, it is also true that the use of DL in epidemiological analysis has a dangerous potential of unethicality, as well as formal problems [67,68]. Again, however, the evaluation of involved expert agents will evaluate such difficulties as things to be solved or huge obstacles for the advancement of the field.

5. Conclusion: Causal Evidence Is Not A Result, But A *Process*

I have made an overall reply to the main critics of deep learning (and machine learning) as a reliable epistemic tool. The basic arguments of Judea Pearl have been analyzed using real examples of DL, but also by making a more general epistemic and philosophical analysis. Following the ideas of Schölkopf [69], I consider that this debate has several difficulties of a conceptual nature: how we conceived causality, and then how statistical inferences became stronger once databases increases, allowing again to define such data connections or patterns as causal relations. The important aspect here is to admit that big data analysis are facing independent and identically distributed problems (IID). Here it is fundamental to see that variables or mechanisms are giving support to data, beyond blind convergences. Consequently, causality would be connected through data to the physical mechanisms that generate statistical dependencies, showing the role of invariance [70].

The systemic nature of knowledge, also situated and even biased, has been pointed out as the fundamental aspect of a new algorithmic era for the advance of knowledge using DL tools. If formal systems have structural dead-ends like incompleteness, the bio-inspired path to machine learning and DL becomes a reliable way [71,72] to improve, one more time, our algorithmic approach to Nature. Finally, thanks to the short case study of epidemiological debates on causality and their use of DL tools, we have seen a real implementation case of such an epistemic mechanism. The advantages of DL for multi-causal analysis using multi-modal data have been explored as well as some possible critics.

References

1. Heisig, J.W. *Philosophers of Nothingness: An Essay on the Kyoto School*; University of Hawai'i Press: Honolulu, HI, USA, 2001.
2. Vallverdú, J. The Situated Nature of Informational Ontologies. In *Theoretical Information Studies*; World Scientific: Singapore, 2019; pp. 353–365.
3. Schroeder, M.J.; Vallverdú, J. Situated phenomenology and biological systems: Eastern and Western synthesis. *Prog. Biophys. Mol. Boil.* **2015**, *119*, 530–537.
4. Vallverdú, J.; Schroeder, M.J. Lessons from culturally contrasted alternative methods of inquiry and styles of comprehension for the new foundations in the study of life. *Prog. Biophys. Mol. Boil.* **2017**, *131*, 463–468.
5. Vallverdú, J. *Bayesians Versus Frequentists: A Philosophical Debate on Statistical Reasoning*; Springer: Berlin/Heidelberg, Germany, 2016.
6. Pearl, J. Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution. *arXiv* **2018**, arXiv:1801.04016.
7. Pearl, J.; Mackenzie, D. *The Book of Why: The New Science of Cause and Effect*; Basic Books: New York, NY, USA, 2018.
8. Gagliardi, F. The Necessity of Machine Learning and Epistemology in the Development of Categorization Theories: A Case Study in Prototype-Exemplar Debate. *Comput. Vis.* **2009**, *5883*, 182–191.
9. Everitt, T.; Kumar, R.; Krakovna, V.; Legg, S. Modeling AGI Safety Frameworks with Causal Influence Diagrams. *arXiv* **2019**, arXiv:1906.08663.
10. Gal, Y. Uncertainty in Deep Learning. Ph.D. Thesis, University of Cambridge, Cambridge, UK, 2017.
11. Kendall, A.G. *Basic Books Geometry and Uncertainty in Deep Learning for Computer Vision*; University of Cambridge: Cambridge, UK, 2017.
12. Kendall, A.; Gal, Y. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? *arXiv* **2017**, arxiv:1703.04977.
13. Piironen, J.; Vehtari, A. Comparison of Bayesian predictive methods for model selection. *Stat. Comput.* **2017**, *27*, 711–735.
14. Polson, N.G.; Sokolov, V. Deep Learning: A Bayesian Perspective. *Bayesian Anal.* **2017**, *12*, 1275–1304.
15. Bengio, Y.; Lecun, Y. *Scaling Learning Algorithms Towards AI To Appear in "Large-Scale Kernel Machines"*; George Mason University: Fairfax, VA, USA, 2017.
16. Cunningham, J.P.; Yu, B.M. Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* **2014**, *17*, 1500–1509.
17. Bengio, Y.; Bengio, S. Taking on the curse of dimensionality in joint distributions using neural networks. *IEEE Trans. Neural Netw.* **2000**, *11*, 550–557.
18. Ben-David, S.; Hrubeš, P.; Moran, S.; Shpilka, A.; Yehudayoff, A. Learnability can be undecidable. *Nat. Mach. Intell.* **2019**, *1*, 44–48.
19. Anagnostopoulos, C.; Ntarladimas, Y.; Hadjiefthymiades, S. Situational computing: An innovative architecture with imprecise reasoning. J. Syst. Softw. 2007, 80, 1993–2014.
20. Raghavan, S.; Mooney, R.J. Bayesian Abductive Logic Programs. 2010. Available online: https://www.ijcai.org/Proceedings/11/Papers/492.pdf (accessed on 6 February 2020).
21. Bergadano, F.; Cutello, V.; Gunetti, D. Abduction in Machine Learning. In *Abductive Reasoning and Learning*; Springer Science and Business Media: Berlin/Heidelberg, Germany, 2000; pp. 197–229.
22. Bergadano, F.; Besnard, P. *Abduction and Induction Based on Non-Monotonic Reasoning*; Springer Science and Business Medi: Berlin/Heidelberg, Germany, 1995; pp. 105–118.
23. Mooney, R.J. *Integrating Abduction and Induction in Machine Learning*; Springer: Berlin/Heidelberg, Germany 2013.
24. Vapnik, V.; Izmailov, R. Rethinking statistical learning theory: Learning using statistical invariants. *Mach. Learn.* **2019**, *108*, 381–423.
25. Vapnik, V.; Izmailov, R. Rethinking statistical learning theory: Learning using statistical invariants. *Mach. Learn.* **2019**, *108*, 381–423.
26. Vladimir, V. *Transductive Inference and Semi-Supervised Learning*; MIT Press: Cambridge, MA, USA, 2013.

27. Mao, J.; Gan, C.; Kohli, P.; Tenenbaum, J.B.; Wu, J. The Neuro-Symbolic Concept Learner: Interpreting Scenes, Words, and Sentences From Natural Supervision. *arXiv* **2019**, arxiv:1904.12584.
28. Pearl, J. The algorithmization of counterfactuals. *Ann. Math. Artif. Intell.* **2011**, *61*, 29–39.
29. Lewis, D. Counterfactual Dependence and Time's Arrow. *Noûs* **2006**, *13*, 455–476.
30. Ramachandran, M. A counterfactual analysis of causation. *Mind* **2004**, *106*, 263–277.
31. Vallverdú, J. *Blended Cognition: The Robotic Challenge*; Springer Science and Business Media: Berlin/Heidelberg, Germany, 2019; pp. 3–21.
32. Rzhetsky, A. The Big Mechanism program: Changing how science is done. In Proceedings of the XVIII International Conference Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL'2016), Ershovo, Russia, 11–14 October 2016.
33. Gunning, D & Aha, D.W. DARPA's Explainable Artificial Intelligence (XAI) Program. 2019. Available online: https://doi.org/10.1609/aimag.v40i2.2850 (accessed on 6 February 2020).
34. Casacuberta, D.; Vallverdú, J. E-science and the data deluge. *Philos. Psychol.* **2014**, *27*, 126–140.
35. Calude, C.S.; Longo, G. The Deluge of Spurious Correlations in Big Data. *Found. Sci.* **2017**, *22*, 595–612.
36. Zenil, H.; Kiani, N.A.; Zea, A.A.; Tegnér, J. Causal deconvolution by algorithmic generative models. *Nat. Mach. Intell.* **2019**, *1*, 58–66.
37. Zenil, H.; Kiani, N.A.; Marabita, F.; Deng, Y.; Elias, S.; Schmidt, A.; Ball, G.; Tegnér, J. An Algorithmic Information Calculus for Causal Discovery and Reprogramming Systems. *iScience* **2019**, *19*, 1160–1172.
38. Gustafsson, C.; Vallverdú, J. The Best Model of a Cat Is Several Cats. *Trends Biotechnol.* **2015**, *34*, 207–213.
39. Iqbal, A.; Khan, R.; Karayannis, T. Developing a brain atlas through deep learning. *Nat. Mach. Intell.* **2019**, *1*, 277–287.
40. Bourgin, D.D.; Peterson, J.C.; Reichman, D.; Griffiths, T.L.; Russell, S.J. Cognitive Model Priors for Predicting Human Decisions. *arXiv* **2019**, arXiv:1905.09397.
41. Vallverdu, J. Re-embodying cognition with the same 'biases'? *Int. J. Eng. Futur. Technol.* **2018**, *15*, 23–31.
42. Leukhin, A.; Talanov, M.; Vallverdú, J.; Gafarov, F. Bio-plausible simulation of three monoamine systems to replicate emotional phenomena in a machine. *Biol. Inspired Cogn. Archit.* **2018**, *26*, 166–173.
43. Vallverdú, J.; Talanov, M.; Distefano, S.; Mazzara, M.; Tchitchigin, A.; Nurgaliev, I. A cognitive architecture for the implementation of emotions in computing systems. *Boil. Inspired Cogn. Arch.* **2016**, *15*, 34–40.
44. Taniguchi, H.; Sato, H.; Shirakawa, T. A machine learning model with human cognitive biases capable of learning from small and biased datasets. *Sci. Rep.* **2018**, *8*, 7397, doi:10.1038/s41598-018-25679-z.
45. Lake, B.M.; Salakhutdinov, R.R.; Tenenbaum, J.B. One-shot learning by inverting a compositional causal process. In Proceedings of the Advances in Neural Information Processing Systems 27 (NIPS 2013), Lake Tahoe, NV, USA, 5–10 December 2013.
46. Gianfrancesco, M.A.; Tamang, S.; Yazdany, J.; Schmajuk, G. Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA Intern. Med.* **2018**, *178*, 1544–1547.
47. Kliegr, T.; Bahník, Š.; Fürnkranz, J. A review of possible effects of cognitive biases on interpretation of rule-based machine learning models. *arXiv* **2018**, arXiv:1804.02969.
48. Narendra, T.; Sankaran, A.; Vijaykeerthy, D.; Mani, S. Explaining Deep Learning Models using Causal Inference. *arXiv* **2018**, arXiv:1811.04376.
49. Nauta, M.; Bucur, D.; Seifert, C.; Nauta, M.; Bucur, D.; Seifert, C. Causal Discovery with Attention-Based Convolutional Neural Networks. *Mach. Learn. Knowl. Extr.* **2019**, *1*, 312–340.
50. Ahrens, W.; Pigeot, I. *Handbook of Epidemiology*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2014.
51. Vandenbroucke, J.P.; Broadbent, A.; Pearce, N. Causality and causal inference in epidemiology: The need for a pluralistic approach. *Int. J. Epidemiol.* **2016**, *45*, 1776–1786.
52. Susser, M. *Causal thinking in the health sciences concepts and strategies of epidemiology*. Oxford University Press: Oxford, UK, 1973.
53. Susser, M.; Susser, E. Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology. *Am. J. Public Heal.* **1996**, *86*, 674–677.
54. Krieger, N. Epidemiology and the web of causation: Has anyone seen the spider? *Soc. Sci. Med.* **1994**, *39*, 887–903.
55. Buck, C. Popper's philosophy for epidemiologists. *Int. J. Epidemiol.* **1975**, *4*, 159–168.
56. Gillies, D. *Judea Pearl Causality: Models, Reasoning, and Inference*; Cambridge University Press: Cambridge, UK, 2000.
57. Tucci, R.R. Introduction to Judea Pearl's Do-Calculus. *arXiv* **2013**, arXiv:1305.5506.

58. Greenland, S.; Pearl, J.; Robins, J.M. Causal diagrams for epidemiologic research. *Epidemiology* **1999**, *10*, 37–48.
59. VanderWeele, T.J.; Robins, J.M. Directed Acyclic Graphs, Sufficient Causes, and the Properties of Conditioning on a Common Effect. *Am. J. Epidemiol.* **2007**, *166*, 1096–1104.
60. Ioannidis, J.P. Randomized controlled trials: Often flawed, mostly useless, clearly indispensable: A commentary on Deaton and Cartwright. *Soc. Sci. Med.* **2018**, *210*, 53–56.
61. Ioannidis, J.P.A. The Proposal to Lower P Value Thresholds to .005. *JAMA* **2018**, *319*, 1429–1430.
62. Krauss, A.. Why all randomised controlled trials produce biased results. Annals of medicine, 2018, 50.4: 312-322
63. Shrier, I.; Platt, R.W. Reducing bias through directed acyclic graphs. *BMC Med Res. Methodol.* **2008**, *8*, 70, doi:10.1186/1471-2288-8-70.
64. Doll, R.; Hill, A.B. Smoking and Carcinoma of the Lung. *BMJ* **1950**, *2*, 739–748.
65. Fisher, R.A. Lung Cancer and Cigarettes? *Nature* **1958**, *182*, 108.
66. Bellinger, C.; Jabbar, M.S.M.; Zaïane, O.; Osornio-Vargas, A. A systematic review of data mining and machine learning for air pollution epidemiology. *BMC Public Health* **2017**, *17*, 907, doi:10.1186/s12889-017-4914-3.
67. Weichenthal, S.; Hatzopoulou, M.; Brauer, M. A picture tells a thousand…exposures: Opportunities and challenges of deep learning image analyses in exposure science and environmental epidemiology. *Environ. Int.* **2019**, *122*, 3–10.
68. Kreatsoulas, C.; Subramanian, S. Machine learning in social epidemiology: Learning from experience. *SSM-Popul. Heal.* **2018**, *4*, 347–349.
69. Schölkopf, B. Causality for Machine Learning, arXiv: 1911.10500v2 [cs.LG] 23 Dec 2019. Available online: https://arxiv.org/abs/1911.10500 (accessed on 6 February 2020).
70. Rojas-Carulla, M.; Schölkopf, B.; Turner, R.; Peters, J. Invariant models for causal transfer learning. *J. Mach. Learn. Res.* **2018**, *19*, 1309–1342.
71. Drumond, T.F.; Viéville, T.; Alexandre, F. Bio-inspired Analysis of Deep Learning on Not-So-Big Data Using Data-Prototypes. *Front. Comput. Neurosci.* **2019**, *12*, 100, doi:10.3389/fncom.2018.00100.
72. Charalampous, K.; Gasteratos, A. Bio-inspired deep learning model for object recognition. In Proceedings of the 2013 IEEE International Conference on Imaging Systems and Techniques (IST), Beijing, China, 22–23 October 2013; pp. 51–55.