Journal de la Société Française de Statistique
Vol. 161 No. 1 (2020)

Prometheus unbound or Paradise regained: the concept of Causality in the contemporary AI-Data Science debate

Richard Starmans[1] [2]

Abstract: This essay highlights some aspects, core themes and controversies regarding causality from a historical-philosophical perspective with special attention to their role in the AI-data science debate. Firstly, it outlines the contours of this debate and subsequently addresses the aporia of causality in statistics, AI and the philosophy and science. In view of the prevalent crisis some key themes and controversies are identified, and a frame of reference is proposed, that may clarify historical controversies and the current state of "agreeing to disagree" in science and philosophy. Secondly, the essay highlights the historical scope of the concept, outlines some early perspectives and "key moments", that involved main conceptual shifts. Thirdly, the essay outlines the rise of statistics and its role in attempting to defuse the crises by entering a sort of progressing liaison with causality. Finally, it is shown how research in AI has further shaped the concept and how and why causality is about to play a crucial role in the current quest for responsible, explainable and transparent AI and data science.

Keywords: Artificial Intelligence (AI), Causality, Data Science, Philosophy, Statistics
AMS 2000 subject classifications: 62A01, 68T01, 97R40

1. Prologue: Prometheus and Pandora's dowry

1.1. Romantic Science and the Fragmentation of Knowledge

The English poet Percy Bysshe Shelley (1792-1822) was not merely a highly sensitive romantic lyric, who made the reader part of his deepest turmoil and provided insight into his tormented spirit. Above all, he intended to be a profound "philosophical poet", not shunning heavy metaphysical reflections, epistemic meditations or moral dilemmas. In his posthumously published essay *A Defense of Poetry* (1840), Shelley goes one step further and argues that the art of poetry embodies a higher form of knowledge than just (natural) philosophy and science. The poet attempts to gain insight into the true nature of the phenomena and their underlying (first) principles or, more eloquently put by Faust in Goethe's famous drama, "*daβ ich erkenne was die Welt am innersten zusammenhält, Schau alle Wirkenskraft und Samen und tu nicht mehr in Worten kramen*" ("that I know what holds the world most closely together, see all the power and the seeds, and do not do any more in words") Goethe (1808). As such, this knowledge concerns the "Great Chain of Being", an idea derived from Plato and Plotinos, further developed in medieval philosophy and revitalized by Arthur O. Lovejoy (1873-1962) in his eponymous study Lovejoy (1938). To that end, it is necessary to know the *causes of things*, following the renowned dictum "felix qui potuit rerum cognoscere causas" ("blessed is he who understands the causes of

[1] Utrecht University
E-mail: R.J.C.M.Starmans@uu.nl
[2] Tilburg University

things") put forward by Publius Vergilius Naso in his Georgics Virgil (2009). In the philosophical poets search for total, holistic rather than fragmentary knowledge, there was neither a priori distinction between *explaining* or *understanding* the *outer* and the *inner* world, nor an urge for subordinating impressions and the fullness of human experience to a sterile ideal of objectivity, or a necessity to oppose the "homo mensura" to natural order. Re-narrating, re-creating and re-interpreting ancient myths was unmistakably part of it. Creative urge, contemplation by the gifted or chosen one, imaginative power, evocating and enchanting reality by language, all contributed to what some called a romantic conception of science Berlin and Hardy (1999). This episode in the history of ideas, which has long been undervalued and interpreted pejoratively, but which has undergone a considerable reappraisal in the last 40 years, is of course far from monolithic Starmans (2018b). With some good will even Von Humbolds "Bildungsideal" can be typified as such, but many would agree that the movement found a highlight in Goethe's famous theory of light and color Goethe (1982), which challenged Newtons optics and was published for the first time in English in the same year (1840) Shelley's aforementioned essay was released.

All this played out on the eve of famous 19th century philosophical debates on knowledge and methodology between David Hume-inspired empiricists (Bentham, Stuart Mill, Mach, Pearson) and Auguste Comte-inspired positivists (Durkheim) on the one hand, and neo-Hegelian (Dilthey) or neo-Kantian thinkers (Rickert, Cassirer, Windelband) and hermeneutic philosophers (Schleyermacher) on the other hand. Their heritage includes many famous distinctions that are still being used today: Erklären versus Verstehen (Dilthey), nomothetic versus idiographic (Windelband), empirical / analytic versus interpretative / hermeneutic approaches and of course quantitative versus qualitative research. Despite audacious early attempts of reconciliation by Max Weber, all this would tamper the probabilistic revolution in the second half of the 19th century Starmans (2018d) and more importantly, it would lead to the current schism of a continental and Anglo-Saxon philosophy, that still casts a shadow on many contemporary issues, including *the theme of causality.* Of course, this philosophical debate could only emerge against the background of the rise of modern science in the 19th century, the proliferation of new disciplines, fragmentation of knowledge, spectacular progress in mathematics and physics, recurrent foundational crises in recently emerged disciplines like psychology, sociology and economics and a process of historizing of the worldview Starmans (2011a). And, last but not least, the probabilistic revolution that would greatly affect nearly all aforementioned disciplines, their foundations and methodology Krüger et al. (1981, 1987).

1.2. The myth of Prometheus and the Philosophy of Technology

In 1820, the very same Percy Shelley wrote his famous poem *Prometheus Unbound*, a lyrical drama consisting of four acts, based on the ancient myth in which the immortal Titan son Prometheus evoked the wrath of Zeus by stealing fire from the Olympus and making it available to mankind. He was then sentenced to an eternal punishment, chained to a rock in the Caucasus. An eagle, sent by Zeus ate his re-growing liver every day, until the mortal hero Hercules finally managed to kill the bird of prey and to break the chains of Prometheus. The latter scene is one of the most portrayed and depicted myths, not only in the narrative tradition, but in the entire Western history of art. Shelly's re-creation aptly illustrates the development that the old myth has gone through in about 2500 years. The story has traditionally been based on divergent, fragmentary and partly contradictory sources, going back to Hesiodos' *Theogonia*, the classical play *Prometheus Bound*, written by / attributed to the Attic theater poet Aeschylos and of course Plato's dialogue *Protagoras*. All adapted Prometheus to the spirit of their time or artistic needs and aspirations, and as a result plot and character became more and more complex. Prometheus did not merely figure as a one-dimensional rebel or renegade, who perished because of his "vana curiositas" or "hubris", a lesson and prophylactic warning to lower vessels not to withstand the Supreme God or mother nature (like Sisyphos, Tantalus, Icarus, Phaethon and Orpheus all did for divergent reasons). In fact, Prometheus could grow into a complex character that became an emblem of the human search for knowledge, the (in)possibilities of science, the grandeur and misery of technology in relation to natural order and homo mensura, the "natural" versus the "artificial", that has stood the test of time and is still prominently manifest in the Philosophy of Technology. Given the dramatic plot it is hardly surprising that the pejorative meaning dominated the narrative tradition, especially in the horror and science fiction literature. The former actually started with Percy Shelley's equally renowned sister Mary, who ominously gave her horror novel *Frankenstein* (1818) as a subtitle *The Modern Prometheus*. In the Philosophy of Technology, the myth would also exert an unprecedented influence, but there is a more balanced picture here. Indeed, denunciation of technology is dominant, ranging from Romantic criticism on the Enlightenment and the scientific worldview, expounded in Abrams' classic *The Mirror and the Lamp* Abrams (1953) to Martin Heidegger's pivotal point in thinking about technology *Die Frage der Technik* Heidegger (1954)); ranging from Horkheimer and Adorno's Marxist criticism in *Dialektik der Aufklarung* Horkheimer and Adorno (1972) to Habermas aversion of the glorification of "Zweckrationalität" Habermas (1981). However, the antithesis is well-articulated too. Ranging from high expectations of the evolutionary ethic Herbert Spencer that mankind would reach his eudaimonic completion, resulting into a state where egoism and altruism would converge Spencer (1879) to contemporary trans- and post humanistic utopias Bostrom (2005). And, ranging from Helmut Plessners characterization of man as a "naturally artificial being" Plessner (1928), Donna Haraway's *A Cyborg Manifesto* Haraway (1985) and Andy Clark's *Natural-born Cyborg* Clark (2003) to a more recent reinterpretation of the myth of Prometheus by the French philosopher of technology Bernard Stiegler. The latter argued in many books, essays and interviews that this creation myth depicts man as a poorly endowed, incomplete animal without essence and defining qualities, and he blamed generations of philosophers for failing to understand that it is just technology that compensates for this, rather than being some alienating, negative force. He did this in various parts of his sometimes rather cryptic and until now unfinished magnus opus *La Technique et le Temps*, the first part of which appeared in 1994 Stiegler (1994), but also in a more accessible way in bundled interviews and transcripts of radio interviews Stiegler (2014).

Be that as it may, it is clear that at this very time a new episode is being added to the genealogy of the Prometheus myth. Or rather, the process of retelling, re-creating and reinterpreting the myth of the unleashed Prometheus has entered a new phase. This becomes clear against the background of the AI's data science debate that has held society, politics and science in its grip for more than five years Starmans (2019b). Central to this are a revived striving for a strong AI; building intelligent, autonomous machines with a "mind" that possesses human qualities such as consciousness, emotions, language and morality. Machines that can communicate and interact with human beings, based on our *causal language*. In addition, and closely related but more specifically, there has gradually come an understanding that mankind is becoming increasingly dependent on the omnipresence of data and intelligent, opaque, "black box" or deep learning algorithms. The systems that use these algorithms may determine, monitor, assess, convict man, work in synergy with him, but may also dominate or replace him and this obviously entails many moral problems and invokes fervent arguments for Responsible, Explainable, Transparent, Fair and Socially Aware AI and data science. And, more recently the quest for hybrid AI, where symbolic and subsymbolic AI finally seem to meet. The enormous attention in media, politics and science seems to indicate that Prometheus has now definitively thrown off his chains and that mankind is facing a new crisis, now that he can -at least in principle- create or choose his evolutionary successors, but in a dystopian reading he may also evoke the wrath of the machines, which could evoke a state of transhumanism (cyborgs) or even posthumanism Starmans (2015). It all sheds new light on the long-standing question in the philosophy of technology: does technology become a driving, progressive, all-determining force, whereby man loses his grip and autonomy, or does technology above all remain a human construction, conceivable and controllable by man, his regulations, conventions, guided by conscious and well-understood choices and beliefs of people?

Why and how could Prometheus obtain this emblematic status? Prometheus ("foresight") was an immortal son of Gaia, smart and skillful, whose career had a prosperous beginning. In the battle between the Titans and Zeus he wisely chose the latter side. After his victory, Zeus rewarded Prometheus and entrusted him and his mid-gifted brother Epimetheus with creating living creatures out of clay and water. Epimetheus ("hindsight") was rather reckless and got off to a good start by equipping the animals with all kinds of excellent qualities, such as power, speed, a thick fur, sharp claws, wings, etc. When Prometheus became involved, he noticed that there was very little left for man, making him a weak creature, helpless and heavily dependent on the Gods. This creature had to use all kinds of tricks and technology to compensate for his lack of real qualities. Prometheus tried to help mankind, but Zeus was very reluctant with respect to this initiative. Especially when Prometheus overplayed his hand by starting to disobey and even deceive the gods. For example, when Zeus forced mankind to make an atoning sacrifice and asked for the best part of a slaughtered bull, Prometheus mislead Zeus and made him choose the wrong part that consisted mainly of bones and fat. When Zeus took away the fire from man, Prometheus stole it back, after which his destiny was sealed. But that was not enough for Zeus. In order to punish mankind itself Zeus made Hephaistos, the crippled blacksmith of the Gods, to create Pandora ("the all-endowed"), the first woman. The Gods gave her all sorts of qualities such as beauty, wisdom, curiosity, language proficiency and many more. Zeus gave her a small box as a dowry, containing all plaques imaginable. The box contained the warning that it should never be opened. Because Zeus knew that Prometheus was too smart and suspicious to accept any gift from him, he made Pandora the bride of Epimetheus, knowing that the less intelligent Epimetheus and the curiosity-driven Pandora would jointly open the box. The outcome is known; the box was opened, the plagues came over the world, but with a second opening hope could escape and came over the world too. This abridged and admittedly far from eloquent rendition of the myth will suffice here to highlight a few characteristics.

1.3. Prometheus Unbound or Paradise regained?

First, the creation of man appeared to be a complicated and thorny project, with Prometheus, Epimetheus and Zeus all three involved, with no proper division of labor, no building plan. There was only limited insight into the process and the mechanism of creation, no teleology, and no way to predict the outcome at all. As a result, mankind had no essence and there were no qualities that really defined man. Uncertainty, indeterminacy and imperfection were immanent. It would appear that no one felt overall responsible. All this induced a problematic triangle between Prometheus, mankind and the gods. This becomes clear if we apply the most famous ancient *theory of causality* to the myth: Aristotle's theory of the four causae ("aitiai"). According to Aristotle to understand an entity or phenomenon means to understand its causes: causa materialis, formalis, efficiens and finalis. They correspond with the answers to four questions: What is it made of? What is it that makes it what it is and not something else? Who initiated it? What is it made for? A full-fledged answer to all these questions is a prerequisite to each *scientific explanation*. The combination of causa materialis and formalis resulted in the essence, Aristotle's famous substantial forms. The causa efficiens brings about the effect or change; the form as present in the spirit of causa efficiens is transposed to the product, for the sake of purpose, the causa finalis. Obviously, the link between matter and form in creation of mankind was rather arbitrary here, due to lack of intentionality and lack of a plan or picture in the mind of causa efficiens. As a result, man had no essence, became largely incomprehensible, had no real consciousness, was virtually out of control, had no understanding of the contingencies of being, his roots and destination. So, generally speaking, the whole idea of causal explanation was absent. Both creator and creation were far away from the philosophical poet's knowledge ideal and didn't grasp the great *chain of being* and *the causes of things*. On the one hand Pandora's role as the first female intended to punish humanity may be rather pejorative, on the other hand the first woman at least was named, well defined and equipped with many outstanding qualities, in fact a joint project of the gods, whereas the first man was nameless, incomplete and without essence. All this does not seem to be more than meager comfort. The problematic status of mankind - man and woman - does not diminish, nor does the frightening dowry and its relationship to natural order.

In addition, it must be noted that Prometheus tried to play his role as a mediator between Zeus and mankind, but he was far from successful and in fact there was a fate connection with man. Both received an eternal punishment, for Prometheus there was redemption due to Hercules, for mankind only the second opening of Pandora dowry brought relief. And of course, the possibility and perhaps paradoxically also the assignment - after the liberation of Prometheus - to enter into the reciprocal relationship with technology, to fully exploit it and to search for an essence or at least for the urgently needed, but still missing qualities. The latter then leads to another salient point. The story shows how Prometheus commits itself to the homo mensura or human condition, embodies or exemplifies it so to speak, and is committed to helping humanity, to give mankind qualities that he had to lack because of his imperfect conception, but without which the same human condition was actually defined and recorded. Prometheus considered it a moral imperative to empty human needs and the story increasingly underwent a shift, also in Shelly's portrayal, leaning on a different classical theme. Homo faber is no longer a faustic icon, a renegade that is being destroyed by hubris, or a fatal subject who has to accept the dowry of Pandora with distress. Man should acquire to which he is rightfully entitled, better yet, regain what he once lost or what was taken away from him. This is an old thought in Greek thinking. According to Plato the soul initially had perfect knowledge in the realm of Ideas, but after the “fall into the body” this knowledge was lost and could only be found again and regained through memory or anamnesis. The same naturally applies to the lost kingdom of Atlantis and in optima form in Christian belief after mankind was expelled from the Garden of Eden. The next step is obvious then, the ultimate goal must be to regain paradise. This grand venture is described in the poem Paradise Regained (1671) by the English poet John Milton (1608-1674) in an unsurpassed way. The expressiveness and scope of the story also allow more profane or secular readings and interpretations, e.g. concerning technology. The idea that man can regain his place and status through technology is not present in the old Prometheus myth, but gradually a shift is taking place to a profane reading of the history of salvation, from sin to salvation, whereby man due to technology is reaching a, maybe once already obtained but lost, eudemonic completion, experienced as paradise or a state of self-realization. Against this narrative background, the contemporary AI data science debate can be situated that displays all the characteristics of a technology debate, as will be explained in the following sections. Indeed, many people do question whether it is a good thing that Prometheus is unbound and unchained, considering the problems with his first creation. Especially since, as we stated in the previous section, the AI data science debate shows that now a new episode is being added to the myth with two related aspects: the intended completion of the project of Strong AI and the search for Responsible and Explainable AI.

It will be argued that, since both are not likely to succeed without a proper account of causality, contemporary research on causality will to a large extent be motivated by these issues and its scientific and societal relevance and impact will be increased accordingly.

2. Overview of the essay

In this essay we highlight some aspects, core themes and controversies of causality from a historical-philosophical perspective, with special attention to their role in the AI data science debate. To this aim the essay is roughly structured as follows. Firstly, we will sketch the contours of nowadays AI-data science debate (Section 3) and subsequently we address the aporia of causality in philosophy and science (Section 4). After a short section on the culturally and linguistically inspired approaches to causality (Section 5), it will be argued that a modest conceptual analysis of the concept can be worthwhile, especially since a generally accepted definition of causality is not available - not even within statistics or research methodology, let alone within philosophy (Section 6). Such a conceptual analysis usually comes down to “dragging” the concept within the Philosophical Triangle, built up by the concepts of reality, mind / thought and language / representation, and then analyzing the subtle relationships and interactions between these concepts. In other words, the search for a suitable frame of reference that takes into account both historical and contemporary issues in the burgeoning literature on causality. In view of the prevalent crisis, we start by identifying some key themes and controversies. Here we limit ourselves to formulating seven key questions that may clarify historical controversies and serve as a frame of reference for the current state of “agreeing to disagree” and contemporary debates in science and technology. Many of these questions are firmly rooted in the philosophical literature and will be introduced here informally and briefly. For clarification, they will be related to some specific key moments in the history of causality and especially to general issues in (science) philosophy and AI.

Secondly, we will briefly highlight the historical scope of the concept, outline some early perspectives and "key moments" that concerned main conceptual shifts, all related to the aforementioned seven key questions and the permanent state of crisis (Sections 7 and 8). To this aim, especially the genealogy of philosophical criticism on causality will be addressed as well as some aspects of the prevalent pluralistic view. Thirdly, we outline the rise of statistics and its role in attempting to defuse the crises by entering a sort of progressing liaison with causality (Section 9). However, this has not prevented the situation of today, leaving a seemingly laborious dialogue as a bleak surrogate for unity and cooperation. On the other hand, since nearly all sciences experienced a probabilistic revolution, their approaches to causality are -be it in different ways- typically based on probability and statistics as well. This makes the laborious dialogue a big challenge and no mistake!

Fourthly, and finally we will show how research in AI has further shaped the concept and indicate how and why it is about to play a crucial role in the current AI-data science debate (Section 10). AI or rather the philosophy of AI has pushed the discussion on causality to the next level, and this particularly applies to such divergent fields as knowledge representation and reasoning and machine learning. It plays a key role in the realization of the ambitions of Strong AI aimed at modelling intelligent behavior or mimicking consciousness and the human mind, but in in ethics as well. As already stated, all these concepts dominate international research agendas, political debates and public societal discussions, and causality does play a significant role in it. The epilogue (Section 11) will comprise some challenges for the Philosophy of AI and data science with respect to causality.

3. The contemporary AI-data science technology debate

3.1. The liaison of AI and data

Technological developments with far-reaching consequences for people and society often cast their shadows ahead. Sometimes this manifests itself in speculative philosophical writings, in futurological works or in artistic expressions of (mostly dystopian) science fiction. The feasibility of such a technology does not even have to be proven in order to generate considerable attention in the news media and to provoke public debates. The current interaction between the project of AI on the one hand and the emergence of data science / big data on the other is one of the most eye-catching and controversial technological innovations of the last decades. This symbiosis does not reveal the fact that both disciplines have their own roots, tradition and orientation. For example, according to the apologists of data science, a data revolution has taken place in recent years, we are witnessing a data explosion or, better still, a data tsunami and we are unmistakably living in a dataficated world. The omnipresence and diversity of data, the speed with which they become available and their almost instantaneous analysis, nourish the idea that reality in all its complexity is almost entirely captured, encoded and codified in data. The primacy lies with "the new gold"; raw, uninterpreted data, which has pushed the conceptually richer and more philosophically interesting concept of information into the background Starmans (2019b). Some go one step further and claim that a completely new conception of knowledge and science is created, in which theory formation, causality, semantics and interpretation become obsolete Anderson (2008). The AI's project, on the other hand is much older than the current preoccupations with data science. Even if we ignore classical initiatives and anticipations of Raymond Lull (13th century), Bacon, Descartes, Pascal, Hobbes, Leibnitz (all 17th century), Boole and Babbage (19th century), AI's has a respectable tradition that goes back until mid-last century. It all started in the 1940s when Warren McCulloch and Walter Pitts introduced their renowned neural network, a computational model that referred to "the ideas immanent in nervous activity" McCulloch and Pitts (1943). In 1950, Alan Turing wrote his now classic philosophical essay on the possibility of artificial intelligence and formulated a thought experiment that would lead to the Turing Test Turing (1950). In the same year, Claude Shannon published his groundbreaking article on computer chess in Philosophical Magazine. He described his work as "perhaps of no practical importance", but "of theoretical interest" and that could possibly offer a solution for "problems of a similar nature and of greater significance" Shannon (1950). Shannon then made some suggestions (machine translation, planning, decision-making, automatic deduction) that would set the agenda for the legendary Dartmouth conference in 1956, which is generally regarded as the beginning of AI as a scientific discipline. It was John McCarthy who organized this conference and coined the term AI. Because of the far-reaching ambition to automate human intelligence and to develop machines that can think and reason, the profession has always managed to attract the attention of philosophers. Even the great AI pioneers themselves were aware of the philosophical implications of their work from the very start, which appealed to the imagination of many and effortlessly found their way into the public debate. AI and philosophy of AI have always been two sides of the same coin.

Anyway, the aforementioned symbiosis is also historically salient, because AI has long been associated with models of the human mind, with knowledge systems that symbolically represent at a high level of knowledge and reason with it, and with smart (heuristic) algorithms. AI has traditionally been opposed to brute force solutions; the complete search of large data structures based on computing power. The latter aspect of big storage and fast retrieval is precisely a characteristic of big data. Current advances in AI in general and machine learning in particular are to a large extent "data-driven". A further nuance is not needed here, but we note that it was for the main part the hardware (miniaturization, better chips, storage) that brought big data and AI together and moved them forward. Due to the cross-pollination with big data, the high expectations of AI are coming true and more and more promises seem to be fulfilled. In the meantime, the dependence on data and smart algorithms has increased considerably in society; ubiquitous computing, ambient intelligence, intelligent systems and social robots are no longer buzzwords. The same applies to autonomous weapons, drones, self-driving cars and the Internet of Things, where millions of electronic devices can communicate, decide and act without human intervention. Because of all this, debates about the social and ethical implications of AI and data science have recently gained momentum. The attention seems without precedent and manifests itself in excessive attention in news media, but also in international research projects and numerous popular, both utopian and dystopian publications, often aimed at a wide audience. Obviously, the question is whether there is much new under the sun in the light of the history of technology. Is it a renewed interest in long-standing philosophical issues or is there indeed a new testing ground for ethics? We highlight some of the characteristics of the current debate here.

3.2. The public AI - data science debate

Since a detailed explanation of the AI's data science debate is too far-reaching here, we limit ourselves to a few recent publications by opinion leaders, that will make the contours of the problem visible and which are relevant from a philosophical perspective. Admittedly, the choice is subjective, but is based on the fact that all works were written by experts or at least experienced workers in the field, acquainted with the topics, rather than by outsiders (celebrities, politicians, historians, philosophers) who also play their part in the public debate. In 2013, Viktor Mayer-Schönberger (1966), professor of Internet Governance and Regulation at Oxford, together with British science journalist Kenneth Cukier (1968) published *Big Data: a Revolution that will transform how we live, work and think* Mayer-Schönberger and Cukier (2013). The book became an international bestseller, had many reprints and has since been translated into 16 languages. In 2016, Cathy O'Neil (1972), an American mathematician and former analyst on Wall Street, published the book: *Weapons of Math Destruction by the. How Big Data Increases Inequality and Threatens Democracy* (O'Neil, 2016), that also effortlessly found its way to a wide audience. It turned the author, who describes herself as data skeptic, blogger and "loud mouth" into a globally renowned speaker. The Swedish philosopher, futurologist and forerunner of transhumanism Nick Bostrom (1973) has held that last status for many years already. His bulky *Superintelligence Paths, dangers and strategies,* appeared in 2014; in the book he builds on ideas from, among others, the statistician Irving J. Good and futurologist Ray Kurzweil on the approaching singularity; a stubborn concept with many connotations, which here refers to the moment when artificial intelligence becomes "infinite" through exponential growth (or at least understood by humans) and the human being absorbed in this total intelligence that penetrates the universe Bostrom (2014). The titles of the three bestsellers speak for themselves. Perhaps less spectacular, but by no means less influential is *The Second Machine Age; work, progress and prosperity in a time of brilliant technologies* written by Erik Brynjolfsson and Andrew McAfee Brynjolfsson and McAfee (rton), both associated with MIT Sloan School of Management. Whereas the First Machine Age, or Industrial Revolution, lead to a redistribution of labor in which people and machines are complementary, the Second Machine Age is, according to both, much more erratic and will have a more profound effect on the global economy (Brynjolfsson, 2015). Because it may be rather perilous to have one or two authors consider this complex domain in an objective and complete way, we should mention a well-known initiative by John Brockman. In 2015 his *What to think about machines that think; Today's leading thinkers on the age of Machine Intelligence*, Brockman introduces almost 200 scientists, philosophers and intellectuals, who express their views in very short essays. It offers a wonderful sample of concrete facts and fiction, opportunities and threats, admonitions and caveats, expectations and views Brockman (2015). Since this essay is all about causality and the ambitions of AI-data science, we finish this list with the monograph *The book of Why; the new Science of Cause and Effect* (2018), written by Judea Pearl and Dana Mackenzie. Here Pearl continues his "campaign" against statistics that he started in Pearl (2000) and sketches how the problems with machine learning could be solved and how humanlike communication is possible by using his well-known causal graphs and do-calculus. In his more recent short paper *Seven pillars of causality in relation to machine learning*, published in the Transactions of ACM Pearl (2019) the author explains how a three-layer causal hierarchy (association, intervention an counterfactual) and several tools, including do-calculus, algorithmizing of counterfactuals, and causal discovery are corresponding with “seven cognitive tasks” and according to the author they are necessary steps in realizing the ambitions of Strong AI.

3.3. The technological knowledge domain

The aforementioned works can be viewed in the light of some known features of the technological knowledge domain, which are very briefly reviewed here Starmans (2018d). *The first characteristic* concerns the feverish pace at which technological progress often manifests itself, not as a brave application of crystallized and tested scientific theories, but as an emergent and sometimes autonomous process, in a force field in which the relationship between science and technology is reciprocal and many actors with different interests or responsibilities claim their role; stakeholders, often none of whom can fully oversee the consequences. A precautionary principle and manufacturability are problematic in advance. As a result of all this, technology debates are seldomly purely academic, but public. Success or failure may depend on public opinion, political interests, ideology, religious fate and not on scientific or purely rational argument. A *second characteristic* is just as canonical and is consistent with the previous one. New technologies often cast their shadow ahead. They often already have an impact on society and the image of mankind before the technology is actually established or its feasibility is demonstrated. Sometimes this is thematized in the narrative tradition of science fiction, sometimes in speculative or futurological writings, in which not seldomly even old philosophical problems are thematized and scrutinized. With some good will it can be stated that a considerable part of the philosophy of the AI’s still fits in with this futurological / science fiction tradition. The philosopher of technology can hardly ignore these anticipations; they form an essential part of his field of work and therefore the philosopher of technology finds a natural ally in the futurologist. This is particularly manifest in Bostrom (2014), who follows in the footsteps of illustrious predecessors such as Hans Moravec and the aforementioned Andy Clark, Donna Haraway and Ray Kurzweil. The *third characteristic* mainly characterizes the works of Mayer and O’ Neill. In this characteristic, philosophers of technology will recognize the utopia-dystopia contradiction, the ancient blessing-of-curse dichotomy, the state of salvation versus the menace, which is detectable in almost all major technological breakthroughs or innovations, as already indicated in this essays prologue. Mayer-Schönberger and Cukier (2013) sketches the contours of a dataficated world, in which not only unprecedented opportunities for science and commerce are offered, but people can also achieve ultimate self-realization in their personal lives and thereby achieve a kind of eudemonic perfection that does think of the limitless progress optimism of 19th century ethicists, such as Herbert Spencer after the emergence of the theory of evolution. There is little room for doubt and nuance. Mutatis mutandis, this applies to O’Neil’s book, which builds up an equally passionate antithesis. O’Neil worked for a hedge fund for several years and said she lost faith in the mathematical models, which she says are blindly followed. The book is partly a personal account and contains a list of abuses in society which, according to the author, are reinforced or maintained by mathematical models, invariably pejoratively referred to as “WMDs”. For example, the author explains how various emancipatory measures devised by the Obama administration were counterproductive, how recruitment and selection algorithms or mortgage lending further push the poor and the disadvantaged into the defensive. Of course, Wallstreet is also covered in a cynical moral sketch, which fits in seamlessly with the image that is sketched in blockbusters such as Oliver Stone's *Wallstreet* (1987) and more recently Martins Scorsese's *The Wolf of Wallstreet* (2013). Brynjolfsson takes a more nuanced approach and takes a middle position in this respect between Bostrom and O'Neil. In any case, O'Neil's indictment evokes another classical theme, which could be considered *a fourth characteristic*: the alleged often postulated value freedom of science in general and of mathematics in particular. The Dutch mathematician and politician Alexander Rinnooy Kan, for example, clearly states that both pure and applied mathematics have no moral dimension and that bottlenecks only concern the application of models and the chosen goals. "The mathematical model is ethically neutral. If it falls into bad hands, then the model cannot be blamed. There is nothing in the mathematical model that could make it avoidable. There is also nothing in the model that makes it inevitable". The author states that it is worthwhile to better map out "where, when and why statisticians / mathematicians regularly end up in difficult ethical waterways" (Rinnooy Kan, 2012, translation R. S.). Although O'Neil hardly touches on the historical and knowledge-theoretical aspects, she would hardly be inclined to agree to all this. According to her, allegedly objective, value-free techniques appear to lead to morally debatable actions, not intended by vicious characters, but possibly by people of good will. A classical distinction in the philosophy of science manifests itself here; an analytical-empirical conception of science, which aims at an objective, value-free description of reality and which should be considered primarily from an internalist point of view versus a conception of science that actually changes reality or serves to promote certain interests and, above all, requires an externalist approach, in which the economic, cultural or political context plays a role and social criticism seems immanent. This value-driven societal-critical orientation is to be considered a *fourth characteristic of technology debates*. Philosophers from the Frankfurter Schule, including Theodor Adorno, Max Horkheimer and Jürgen Habermas laid the foundation for this position. One can also think of the Soviet historian Boris Hessen, who criticized Newton, because he believed that his classical mechanics was primarily intended to represent the interests of the bourgeois / reigning class. His line or argument was in full accordance with Marx' dialectical materialism. More recent work would include the early sociological studies of Bruno Latour, the Strong Program of Barry Barnes and David Floor, more recently advocators of Science Studies and of course adherents to social constructivism such as the physicist and sociologist A.R. Pickering. In his *History of particle physics: a sociological analysis* from 1983 the first contours of a social-constructivist paradigm emerge which was actually a prelude to his well-known *Constructing Quarks: A Sociological History of Particle Physics* (1984).These all emphasize, in their own way, the social structure of science, the group interactions, institutions and conventions, with which science becomes an activity that is not essentially different from many other human activities. The historian of Science and physicist Peter Galison also emphasized the role of social structures, conventions in determining how experiments are conducted, although he strongly rejects social constructivism, adheres to scientific realism, and also rejects aforementioned externalist perspectives, that study science without detailed domain knowledge, including actual developments and substantial changes in the field. On the contrary, Galison does not shy away from technical and sometimes meticulous details in his historical case studies. Anyway, many of all these approaches have in common that science, society and criticism on society are closely connected, making the idea of a value-free science inconceivable.

Finally, *a fifth characteristic* concerns the underlying concept of man. Contemporary philosophers of Technology sometimes fall back on the German philosopher and founder of anthropology Helmuth Plessner (1892-1985), who developed a biology-inspired vision of homo Faber and states that man is naturally artificial. He therefore does not regard this as a natural, unchangeable and inalienable characteristic of his own, but neither does it give priority to a purely evolutionary interpretation of man. As a philosophical anthropologist, he seeks to define the essence of man in the interaction with invariably changing technology, in which man is constantly rediscovering and redesigning himself. In the current philosophy of technology debate, which is strongly dominated by philosophical anthropologists and ethicists, Plessner's vision is widespread and many refer to it or build on it. Transhumanist strive, Posthumanism and Bernard Stieglers approach have been mentioned in the Prologue, already. Of course, this issue has been dominant in AI for the very start and it dominates the current debate as well, but this aspect will not be dealt with extensively in this essay.

4. The aporia of causality

The concept of causality has played an important, though controversial role in the history of ideas, which continues to this very day. The preoccupations with causality concern both *philosophy* (metaphysics, ethics, epistemology and philosophy of science) and *research methodology* (statistics, research design, modelling), but also come to the fore in the foundations of the individual sciences in many different ways. The notion of causality will be used here straightforwardly as an umbrella term, overarching cause and effect relations, the (metaphysical) *principle* of causality, but also *causation* as a *process*, causal factors, connections and links. It manifests itself in a multitude of appearances, including causal processes and mechanisms, causal forces and powers / dispositions, causal statements and arguments, causal theories and models, causal reasoning and inference Starmans (2018f). In fact, these appearances are often based on different approaches or conceptualisations of causality. The list, which is neither mutually exclusive nor totally exhaustive, can easily be expanded. The manifestations usually take shape against a background of concepts such as logical necessity, physical necessity or determinism and of "dual", equally tricky and thorny notions like contingency, coincidence, probability, chance, uncertainty, indeterminacy, free will and moral responsibility. All these concepts constitute or "build" the concept of causality, form a contrast with it or can be associated with it historically and philosophically.

Anyone who goes through the same history of ideas from the very beginning of pre-Socratic natural philosophers up to the era of data science and AI will inevitably identify causality as a persistent and *complex problem area*. First and foremost, causality has traditionally shown many different faces and appearances; it went through a long genealogy with substantial *conceptual shifts*, thus seemingly being in an almost permanent state of crisis. Among other things, the concept has been regularly eliminated from the scientific language, but it has also been rehabilitated, if not *reinvented* several times, by philosophers who all committed an intellectual father murder by breaking radically with tradition Starmans (2018b). Even today, causality evokes entirely different points of view. For some, the concept is utterly obsolete and outdated. In accordance with classical philosophical criticism from renowned thinkers such as Ernst Mach, Karl Pearson, Bertrand Russell and Paul and Patricia Churchland, criticisers regard causality as an improperly cherished relic from a bygone era. In the wake of Google's research director Peter Norvig, some contemporary sceptics prefer to rely on "the unreasonable effectiveness of data" Halevy et al.

(2009) or conclude that "causality has definitively been pushed away from its pedestal as a primary source or meaning" Mayer-Schönberger and Cukier (2013). Some even proclaim "the End of Theory", contending that "correlation supersedes causation" and that causal models, whether explanatory or interpretative, are obsolete in epistemology Anderson (2008) or they establish their hope for "the Master Algorithm" Domingus (2016) or "the Deep Learning Revolution" Sejnowski (2018). Others, on the other hand, regard causality as "the most immediate and vital element of the world" Mumford and Anjum (2013) "the cement of the universe" Mackie (1980). They try to identify the "causal structure of the world" (Salmon, 1994) or they characterize causal relationships as "the fundamental building blocks both of reality and of human understanding of that reality" Pearl (2000). In doing so, some of them do seem to confess themselves unmistakably to a contemporary form of metaphysical realism. No doubt AI scientist and Turing prize laureate Judea Pearl is currently the most prominent advocate of causality. In his aforementioned *The Book of Why: the New Science of Cause and Effect* (2018), he states outspokenly that a causal revolution is taking place, which will conquer the world, accompanied by an associated "causal mathematical language" Pearl and MacKenzie (2018).

Obviously, there is no shortage of big words and as is often the case, especially in bestselling books aimed at wide audiences, moderate and nuanced positions are somewhat underexposed when powerful metaphors and rhetorical violence are present in excess, and pathos and exaggeration are not shunned. However, there is an even more vital aspect of the prevailing crisis, i.e. the fact that even those who embrace the concept of causality today are in a permanent state of "*agreeing to disagree*" rather than striving for unification or at least to some extent for "*unity in diversity*".

On the one hand, causality is unmistakably "en vogue" Williamson (2009). Today the literature is ample, in epistemology it has pushed more fundamental notions such as truth and validity into the background Starmans (2018f). In the Philosophy of Science, the interactions between causality, scientific explanations and laws / lawlike statements gain much attention, not to mention specific topics like causal realism, mental causation, the causal theory of reference and many more that are widely studied. The same applies to the more formal philosophical literature, such as Glymour (2001). It is often postulated that the need for causality in data science and AI is now stronger than ever Pearl and MacKenzie (2018), van der Laan and Rose (2011, 2018) and A. and Robins (2019). As mentioned, the *public AI-data science debate* increasingly demands Explainable, Responsible and Socially Aware AI and data science (transparency, accountability, ethical awareness, value alignment, et cetera) which does seem to be illusory without a proper account of causality. All these requirements dominate AI Manifestos, research agendas and political policy outlines, embraced by a variety of advocators, including CEO's of leading companies, such as Google's Sundar Pichai, scientific bodies such as the Association for Computing Machinery (ACM), the UN, EU and world leaders and influencers in the political, economic or religious area.

On the other hand, paradoxically, there is little cross-fertilization or cooperation between the different formal approaches to causality, and sometimes mutual misunderstanding and neglect seem immanent and deeply ingrained Starmans (2019b). As a result, and despite considerable technical progress in the separate fields, the burgeoning recent literature on causality has only modestly influenced *actual research practice*, (epidemiology, statistics, data analysis) and – more generally – *research methodology*, let alone the public AI-data science debate. Recent attempts to change the status quo, such as *Causality: Philosophical Theory meets Scientific Practice* Illari and Russo (2016) and *The Oxford Handbook of Causal Reasoning* Waldmann (2017), only reconfirm and reinforce the impasse in an uncomfortable way. For those who need additional arguments, we refer to Anjum and Mumford (2018) and a review of the book by Clark Glymour Glymour (2019). First, many of the aforementioned studies especially disclose the gap between the "epistemological" and "methodological" level, also described in Starmans (2018f). Secondly, and even worse, those who restrict themselves to probabilistic approaches to causality and to the domain of mathematical statistics in particular Burgess and Thompson (2015), A. and Robins (2019), Pearl and MacKenzie (2018), van der Laan and Rose (2011, 2018) will observe no different picture. All in all, the historical crisis appears to be continuing. The *proclamation of pluralism* as a last resort or escape route then becomes almost inevitable, and any *pursuit of unification or unity* seems illusory. Understanding the theme as a "crisis of causality" or even identifying *an aporia* or irresolvable internal contradiction does not even seem to be an awkward choice of tone.

What is left is a plethora of approaches (not mutually exclusive or totally exhaustive) that all try to grasp the "essence" of causality: regularity theories, counterfactuals, interventionist approaches, dispositional-, actor- and process-oriented approaches, mechanistic visions, difference-making methods, potential outcomes, instrumental variables, Bayesianism, Pearl's do-calculus, information-theoretic approaches, et cetera. All aforementioned studies are based on (combinations of) these approaches. *For those who want to solve the problems of the AI-data science debate by relying on existing literature on causality this is not a convenient start.*

5. Culture, Language and Common Sense

In this section we will address another aspect of the problem with causality. We will take a different stance than the one in the previous section and pay attention to three aspects of causality that are a little underexposed in the current data science AI debate: the general cultural-historical dimension of causality, the linguistic turn and finally causality as a primitive / common sense notion.

The concept of causality may still stir the sciences, but in fact it has always been omnipresent in natural philosophy, metaphysics, theology, even in pre-scientific times, because it is part of everyday experiences, discourse, ordinary language, rituals, customs and use. Without exaggeration, it can be said that causality has an impressive cultural-historical reach. Those interested in causal allusions within the narrative tradition (ranging from ancient myths and folk tales to Dostoyevsky's *Guilt and Penance*, Musils *Mann ohne Eigenschaften* (*Man without qualities*) and Nabokov's *Lolita*), in sacral religious texts, but also in quantum mechanics, genetics and psychoanalysis can go to the unsurpassed study *A Cultural History of Causality: Science, Murder Novels and Systems of Thought* (2014) by the American historian Stephen Kern Kern (2014). Roughly put, the author analyses the evolution of thinking about causality and tries to fathom the various theories from literature and, more specifically, from crime literature. Andre Gide's *Lafcadio's Adventures* (1914), in which the main character commits a murder for the sole purpose of committing a murder for no reason, is one of the starting points. Of course, Raskolnikov (*Guilt and Penance*), Th/'er/'ese Raquin (from Emile Zola's novel of the same name) and Hannibal Lector (*The silence of the lambs*) are also discussed. In addition to Gide's "motiveless motives", many conceptions and connotations of causality are discussed. The course of cultural or, more specifically, literary development that Kern outlines, thus appears to be one of the pillars of contemporary pluralism, which makes any attempt at compulsive or naïve unification fail in advance. It could well raise the question whether an axiomatisation of causality would be something to strive for in the first place. In this respect Kern's linguistic-philosophical chapter is particularly relevant; here causal factors of crime are sought not so much in deviant morals, genetic dispositions or mental aberrations, but in problems in and with language itself. For example, a character from Samuel Becketts *Molloy* explains a murder of a man from problems with or within the language: "either I didn't understand a word, or he didn't understand a word I said." Many other sources could be mentioned, but keeping in mind this essays Prologue, here we refer to Frederick Burwicks paper *The Language of Causality in Prometheus Unbound*, published in the Keats and Shelley Journal of 1982 Burwick (1982) This preoccupation with language brings us to the next, already announced aspect.

Despite age-old concern of erudite thinkers with causality, it seems primarily an intuitive concept, a common sense notion or a "natural" category (in the Aristotelian or Kantian sense of the word), which we constantly, directly and sometimes almost instantaneously use as intelligent organisms to orient ourselves in reality, to adapt to the eventualities and capriciousness of life and to uphold in the struggle for life. But also, to *explain* our experiences and to *understand* ourselves, our situation and the contingencies of the human condition. As a result, causality is constantly visible in everyday language. This happens, for example, explicitly with different conjunctions (because, due to, therefore, so, though, nonetheless, but, however, et cetera) and numerous "causal" verbs (cause, induce, lead to, trigger, generate, influence, produce, etc.) All transitive verbs are implicitly causal. A transitive verb's natural interpretation assumes a subject, an actor, or an abstract entity that effects a change in an object or reality, e.g. "Jan eats the sandwich." The corresponding question is also causal. Who eats the sandwich? Answer: the *entity / actor* who commits an *intervention* and initiates a *mechanism* whereby the sandwich gets involved in a *process* of being eaten, and thereby undergoes a *change* and eventually reaches the final state of "being eaten" and "no longer being a sandwich" apparently a *visible difference* from the original state of "not yet eaten" and "still a sandwich." Transitive verbs lead to causal question sentences. That is why many scientific and everyday questions are also *intrinsically causal* Starmans (2018f). In view of the omnipresence of causality in human actions and the very diverse anchoring in language, a linguistic searchlight on causality seems necessary to get a full picture of it. This linguistic perspective is reflected in the philosophy of ordinary language ("meaning-is-use", the language games of Wittgenstein) and the theory of speech acts, but also in the so-called informal logic, argumentation theory, critical thinking movement and discourse analysis, which have all taken off in the last forty years. In this linguistic turn dealing and understanding causality is not stipulating a heavy metaphysical concept, a logical framework or a formal statistical account, deprived from a natural language form. Natural language is not intrinsically vague, unreliable and error-prone like Russell, the early Wittgenstein, Carnap and other logical-positivist thought; it is not only the starting point of the analysis, but the level at which we should conduct our analysis. *The only way to grasp causality, to meaningful use and understand it is to be engaged in a language game, a dialogue where two or more interlocutors are trying to resolve a dispute, to make a decision or negotiate. According to the rules of that specific language game they make claims, arguments and counterarguments, shifting the burden* *of proof, making commitments and change the world by performing speech acts*. Especially in argumentation theory, the study of “topoi”, initiated by Aristotle about 2300 years ago has been pushed to the next level by scrutinizing argumentation schemes, many of which are explicitly or implicitly causal. An important source that also takes a historical-philosophical stance, is the paper *Argumentation Schemes. History, Classifications and Computational Applications,* written by Fabrizio Macagno, Douglas Walton and Chris Reed Macagno et al. (2017). The authors show how argumentation schemes evolved and can be classified today, based on classic and contemporary insights and how they can be “organized in a modular way to describe natural arguments or produce complex arguments”. There are many applications varying from rhetoric and law to AI, decision making and argument mining. So far, this linguistic turn sometimes plays a subordinate role in the current AI data science debate Starmans (2018f). Since causality is in this debate all about explanation, persuasion, accountability and justification, it appears to be highly significant in the current debate on Explainable and Responsible AI. As already indicated some researchers claim that the ideals of strong AI can only be achieved through a dialogue between people and a computer equipped with a “human” notion of causality Pearl (2019). If they are correct, this linguistic approach can hardly be overlooked. In addition, this so-called linguistic turn is, above all, a down-to-earth approach that forms a beneficial antidote to the sometimes heavy metaphysical or knowledge-theoretical approaches to causality.

A third consideration is perhaps even more down-to-earth. Given that causality is above all an intuitive notion, there is no reason to relate it exclusively to high-level cognitive functions such as language. Causal reasoning or rather recognizing cause-effect relationships and causal processes is a crucial aspect of any relevant conceptualisation of intelligence and just as language use and symbol processing are not necessary conditions for intelligence, they are also not necessary for causality. Intelligence as a graduate concept can be understood perfectly low-level on the basis of principles of direct perception Franklin (2014), emergence, situatedness and context dependence Brooks (1991) and that applies equally to specific aspects of causality; as a common sense principle, qualitative, directly given and to be understood, whether or not classified as “evolutionary successful” Starmans (2019b). Also lower forms of life are equipped with it. *In fact, all these considerations do suggest that neither any naïve attempt for unification, nor simply highjacking the concept whether in metaphysics, statistics or in AI, seems the right way to proceed.* Of course, the idea that sometimes causality is to be considered a primitive, self-explanatory or self-evident concept, that need not be further analysed stems from a long philosophical tradition. For example, it is in full accordance with common sense philosophy, that started with Aristotle and had many renowned advocators, including such divergent thinkers as Thomas Reid, George E. Moore, many pragmatist philosophers, and as will be explained, some AI-researchers as well. What’s more, many highly developed engineering disciplines regularly use a common-sense notion of causality, basically a self-evident or self-explanatory, perhaps primitive concept that needs no further formal analysis or theory. Examples can be found in engineering, design science Wieringa (2015), in rapidly expanding disciplines like process mining and even in machine learning Starmans (2019a).

6. “Agreeing to disagree”

Obviously, the previous section only increased the plurality of opinions regarding causality we identified before. Of course, the borders between the narrative tradition, metaphysics, statistics and technical sciences or application domains are not necessarily fixed and demarcated. A recent contribution of Chambaz, Drouet and Thalabard shows this Chambaz et al. (2014). In their paper *Causality, a Trialogue* they pay a tribute to the famous conversation between French philosophers Jean d’Alembert and Denis Diderot, who famously endeavored to restore the old ideal of complete knowledge with their Encyclopedia, a project that started with Cicero and Seneca in the Roman era and that has been continued to this very day with Wikipedia Starmans (2011b). By using this literary genre of a dramatic play as a language game the authors of *Causality: a Trialogue* actually step into a tradition that ranges from Plato’s dialogues and Berkeley’ *Three Dialogues between Hylas and Philonous* (1713) to Douglas Hofstadter’s *Gödel Escher and Bach* (1979). In this trialogue a philosopher (Drouet), a medical doctor (Thalabard), and a statistician (Chambaz) talk about causality. They discuss the relationships between causality, chance, and statistics, resorting to different dialogue games, using a variety of examples to develop their arguments; varying from the narrative tradition and philosophy to modern science and medicine.

Still, as explained in section 3 a conceptual analysis remains desirable for which we will now identify seven key questions (Q1 to Q7) that appear to be useful in illuminating historical and philosophical controversies and may serve as a reference frame for contemporary debates. It should elucidate some of the current miscommunication and may even be helpful to restore dialogue. Many of these questions are well-known in the literature, having invoked a lot of debates, so we will introduce them here informally and briefly in this section. If necessary, they will be expanded, elaborated and clarified then throughout the essay by relating them to some specific *key moments* in the history of causality or by relating them to *general issues in the philosophy of science*.

6.1. Is causality essentially (meta)physical or mental /epistemical? (Q1)

The physical or mental distinction lies at the very heart of the debate. Is causality a physical or metaphysical phenomenon, existing in the real world, mind-independent and objectively or is the concept largely a product, construct or aspect of the human mind, be it an essential one, indispensable to understand the world, to adapt to it, and to survive? Could it perhaps be both? Or is it just some (evolutionary) epiphenomenon that can / should be accounted for or not? Put differently and bluntly, do we need a physicist, a biologist or a psychologist to deal with it? Or, going one step back, maybe even a metaphysician, since causality often starts with abstraction from everyday experiences, phenomena, and dealing with underlying abstract principles. No doubt, key moments in the history of causality related to this question are the contributions of both David Hume (1711-1776) and Immanuel Kant (1724-1804). Both can be considered as the founding fathers of the view that causality is mental, be it in entirely different ways. Their positions will be addressed in the upcoming sections. Secondly, this key question relates to a longstanding and persistent topic in the philosophy of science, i.e. *the scientific realism debate*. Does reality exist mind-independent? Do we have access to it? Can we perceive it, hypothesize or know it with our senses and cognition? Obviously metaphysical positions such as Mumford and Anjum (2011,

2013) should be considered against this background. Suppose someone wants to / doesn't want to take responsibility for a causal claim, or its negation in a dialogue game; does it refer to a force, process, agent or mechanism in reality or is it just a projection or construct of the mind? Does the associated certainty or uncertainty have to do with his state of knowledge or is this all about necessity for this thesis or antithesis to occur, or is the uncertainty intrinsically present in reality? Of course, the corresponding concept of probability (epistemic versus logical or dispositional) plays a role here. The issue is still highly relevant for virtually any discussion today on data science and AI. Thirdly, and finally the philosophical position of *idealism* is relevant here, associated with Berkeley, Kant and Hegel. Berkeley famously stated that "esse est percipii" ("to be is to be perceived"), which demands an observer for something to be able to exist or being real. Many scientists took an idealist stance, including one of the founding fathers of modern statistics Karl Pearson, who was notoriously anti-causalist, but took an idealist, anti-materialist position as well. He thought that the world we are studying in science doesn't exist independently of the human mind and the ultimate goal of science should be an analysis of the human mind and a classification of its content Pearson (2004).

6.2. *Is causality a token or type phenomenon? (Q2)*

The token-type distinction is classic in epistemology and analytical philosophy, and also appears to apply to the problem of causality. Is causality in principle a specific relationship between unique, single-case phenomena that occur "only once" or are we dealing with examples, instances or realizations of underlying rules, general laws or principles? Does causality only occur at the token level and does it manifest itself as an idiosyncratic and unique event or should it be understood and interpreted at the level of the type as an application or an instantiation of a general rule that can be applied repeatedly? The general rule is then supposed *to explain* its instances, the individual phenomena or "multiple occurrences". Since both approaches assign to causality a role in typical explanatory models in science, it goes without saying that this key question relates to another classic subject in the philosophy of science, the triptych "explanation, causality and laws." Even today the notions of scientific explanations, natural laws, and causality are often studied as highly related or associated. An important moment in the history of causality is, of course, the famous *covering-law model* by Carl Hempel, which is still a milestone in the theory of knowledge and explanation and is re-emerging in a new form in contemporary AI-data science debate Starmans (2019b).

6.3. *Do only individual / concrete entities constitute a causal factor or does causality manifest itself at group level / as a composite or even abstract, theoretical entity? (Q3)*

This briefly formulated question, which must be strictly distinguished from the second key question, requires some clarification, because it is based on various philosophical issues. First, one may wonder whether a cause is concrete and material, whether it is a subject, an animated actor or agent that intentionally produces an effect, or an equally concrete and material inanimate object, whose characteristics, tendencies or dispositions will bring about the effect. However, people interact with their environment, which contains more than just other entities and is structured. In fact, individuals participate in groups, clusters, and aggregated entities that structure or create this environment. Statisticians will recognize data that is generally considered to be hierarchical; they are nested, embedded, aggregated or layered, i.e. have different levels and partly explain variation at the individual level or other levels. Can these (allegedly existing) structures, groups, clusters or aggregates be causally active factors? We can go one step further and acknowledge that many concepts in science and daily life are abstract; they are not empirical notions, but theoretical / multidimensional constructions. Can they also be causally active factors? Different key moments in the history of causality can be associated with this key question. We must confine ourselves to only a few of them. Firstly, the distinction between empirical and theoretical terms used by e.g. Rudolph Carnap and other logical positivists and famously challenged by Willard V.O. Quine in *Two Dogmas of Empiricism*. Emile Durkheim's views on social facts are a second historical moment. Social facts are irreducible to individual phenomena and this approach introduces causality at 'group level' in sociology, in contrast to e.g. Max Weber and later adherents to methodological individualism. Thirdly, as a historical moment connected with this key question, the ecological fallacy of Robinson Robinson (1950) and the resulting multilevel analysis in statistics count as a milestone Starmans (2018b). The idea of causal powers Jacobs (2017) could well be considered at the individual / concrete level, intermediate between subject and object, and an example of potential being, rather than active-being in the Aristotelian sense of the word.

6.4. Is causality essentially qualitative or quantitative? (Q4)

This question also has a long tradition in the history of causality, but here we present it in a simple, straightforward way. Does causality include a (physical) quantity that must be represented on some numerical scale based on the correct units? In other words, should it be operationalized and measured accordingly? Should we, for example, quantify a causal effect as a difference between two (hypothesized) means (as the potential outcome method assumes) or as a (standardized) coefficient in a linear regression? Or should we essentially consider causality as qualitative? It could also be argued that the question is not that crucial since both underly the same experimental logic Tacq (2011). Be that as it may, recognizing that both positions can to a certain extent be confirmed by daily experience, there is at least one specific point of view that states that causality is immediately given, that it is a natural category that can be communicated by language and can only be qualitatively interpreted. Different key moments in the history of causality are related to this fourth key question, ranging from Aristotle's qualitative physics to the current distinction between qualitative and quantitative research, based on the 19$^{\text{th}}$ century methodological debate we referred to in the Prologue. Here we limit ourselves to *The Naïve Physics Manifesto* by Patrick Hayes Hayes (1977), a groundbreaking document in AI that forms the basis of the paradigm of qualitative reasoning within AI today. The ability of humans to understand physical phenomena intuitively and qualitatively is, according to many, crucial in the pursuit of Explainable AI and Strong AI in general. Secondly, we refer to a classical philosophical position on causality, which can be traced back to Thomas Reid's commonsense philosophy from the 18th century. No metrics, no underlying abstract structure or microworld, but something that can be viewed and understood immediately. No gap between the everyday familiar world of phenomena and the underlying postulated "real" reality. In a way many quantitative positions do presuppose this.

6.5. *Is causality deterministic or probabilistic? (Q5)*

The relevance of this question seems undeniable and of course, determinism too has many faces. Paradoxically, since antiquity, causality has always been linked to determinism, physical and logical necessity (Descartes, Spinoza, Hobbes), but nowadays almost all causal theories are probabilistic, that is, they are based on probability theory and statistics. In fact, in Starmans (2018f) it is argued that owing to the probabilistic revolution the concept of causality has been able to make a comeback in science in the first place. There are of course many key moments in the history of causality related to this question. Probabilistic theories about causality now also have a respectable philosophical tradition, dating back to the logical positivist Hans Reichenbach (1891-1953), the statistician Irving J. Good (1916-2009) and the philosopher Patrick Suppes (1922-2014) and have also found their way into modern epistemology Williamson (2009). Furthermore, we referred to the successful commonsense interpretation of causality in engineering; of course, the self-evident approach is typically deterministic, but since there is no underlying theory to be analyzed, we will not discuss the issue in extenso here. Finally, we have to refer here to the "logic versus probability" controversy in the AI, which was mainly shaped by Patrick Hayes *In Defense of Logic* from 1978 Hayes (1978) and Peter Cheeseman's antithesis *In defense of Probability* from 1985 Cheeseman (1985). In fact, Cheeseman claimed that classical probability is sufficient to model virtually all aspects of human reasoning. He scorns the "proliferation of new representation languages with associated inference procedures, all extensions of classical logic or unsound mechanism for reasoning with uncertainty. When it comes to automate reasoning with incomplete or uncertain knowledge, common sense reasoning, mimicking human cognition and ultimately making the project of AI successful, "probability is all that is needed". Obviously, this includes causality, as a fundamental aspect of human cognition as well! The battle is still ongoing and certainly not settled, although some form of convergence is detectable. In Section 10 this will be addressed further.

6.6. *Could reasons be considered as a special kind of causes? (Q6)*

This question simplifies the traditional relationship between "causes" (physical processes) and "reasons" (motivations, intentions, goals, teleology). Historically relevant are of course the teleology of Aristotle ("causa finalis") and Thomas of Aquino, the concept of intentionality of Brentano and Edmund Husserl ("aboutness") and especially the Heidegger-based modern AI criticism of Dreyfuss and Winograd, in which aboutness and intentionality of the human being are raised as a decisive objection to the ideal of Strong AI. According to the critics, computers can only achieve symbol processing and do not deal with intentionality. One can also think of multi-agent systems, in which computers are supposed to possess beliefs, desires and intentions, as attempts to counteract this criticism, which makes the question of whether reasons and motives are causal and should be represented as such, very topical. The problem was discussed in the prologue with respect to the great chain of being, the causes of things, but in the everyday language use the difference is often hard to make, as is also quite obvious in the cultural history of the concept, the philosophical literature on informal logic and argumentation schemes, described briefly in Section 5.

6.7. To represent or not? (Q7)

Obviously, section 5 showed that at the language level causality can be represented explicitly using causal words, partly implicitly using words with subtle causal connotations, and sometimes even completely implicitly, skipping all direct references to causal words, but still embracing the cause effect relations, depending on the type of language game one is enrolled in. At the more formal level, this question is of course at the heart of every science: object language versus meta language, explicit versus implicit, categorematic versus syncategorematic, static versus dynamic, concrete versus continuous, the list is almost endless. If a causal mechanism is to be modeled explicitly, should it be represented at the physical, chemical, biological or even "higher" level? If one acknowledges the language level, should one employ the counterfactual, an old and from from a philosophical perspective complicated notion, but recently successfully used by many statisticians and AI-scientists? One aspect that is crucial for the future of causality in the AI's finds a historically important moment in the 1990 publication of Rodney Brooks *Intelligence without Representation* Brooks (1991), in which he stated that intelligence does not coincide with high-level cognitive functions, symbolic or sub symbolic representations, logic versus probability debates or language. It is low-level, emergent, embodied and emergent. In fact, this aspect has already been discussed above. Recognizing cause-effect relationships and causal processes is a crucial aspect of any relevant definition of intelligence and just as language use, symbol processing and reasoning are not a necessary and sufficient condition for intelligence, causality can to some extent do without these issues too. Intelligence as a graduate concept can be understood very well on a low-level basis from principles of direct perception Franklin (2014), emergence, enactment, situation and context dependence Brooks (1991) and that applies equally to specific, crucial aspects of causality. To what extent can one be a genuine causalist, without explicit (formal) representations? As pointed out in Starmans (2018f) many philosophers didn't actually theorize on the concept as such, but highly influenced the development of causality and its role in philosophy and the sciences, simply because the notion was intertwined with their epistemic of metaphysical theories. Today AI-researcher Judea Pearl is the most well-known advocator who demands explicit representations of causality in accordance with the old ideals of Strong AI.

7. Early Perspectives and Conceptual Shifts

We will now show the relevance of these questions by applying them to classical approaches to causality. As stated above, causality went through a remarkable development process in philosophy and in the sciences; the persistence of the use of the term over the centuries is in sharp contrast to the many conceptual shifts it has undergone. The chameleonic and, above all, context-sensitive nature of the concept becomes apparent when we look at metaphysics and epistemology handbooks until the mid-19th century. Various thinkers such as Lucretius (and other atomists), Chrysippos (and other stoics), Plato, Aristotle, Thomas van Aquino, Bacon, Descartes, Galilei, Spinoza, Hobbes, Newton, Leibniz, Locke, Hume, Kant, Stuart Mill and Charles Sanders Peirce are typically included and considered as heirs or figureheads of thinking about causality. That remarkable unity seems to make some canon formation possible, but more relevant is that many of the aforementioned thinkers defined the concept of causality stipulatively and - as befits good philosophers - committed an intellectual father murder by breaking radically with the tradition or at least hardly continued building on illustrious predecessors. In fact, causality was reinvented time and time again, precisely because each *conceptualization was closely linked to the metaphysical and epistemic views of the thinker in question* Starmans (2018b). When a certain philosophical position came under pressure, the associated concept of causality naturally also fell into the dock. Because theories and points of view in philosophy are rarely falsified and replaced, many mutually conflicting visions have continued to co-exist. This remarkable "evolution" lies at the heart of the seven core controversies and remains visible in many contemporary approaches and perspectives: regularity theories, counterfactuals, interventionist approaches, dispositional-, actor- and process-oriented approaches, mechanistic visions, difference-making methods, potential outcomes, instrumental variables, Bayesianism, et cetera Illari and Russo (2016). As explained in Starmans (2018f) almost all of these contemporary perspectives can be traced back to the aforementioned illustrious thinkers. As such we have identified another pillar of contemporary pluralism and a historical source of criticism and crisis. Just as an illustration let us start with some main early appearances and conceptual shifts since the early Greeks and show how they relate to the seven key questions we outlined in Section 3.

Refraining from the aforementioned pre-scientific period, the narrative tradition, and every day or common-sense use of causality, it makes sense to let thinking about the concept begin with the ancient Greeks. About 600 BC the emergence of philosophy and science took place and a naturalistic turn became manifest. Ionic and Doric natural philosophers and other pre-Socratic thinkers tried to explain the phenomena, understand nature from immanently working forces (Q1) and underlying, typically deterministic (Q5) principles, rather than (solely) from divine powers (Q6). What happened was a quest for abstract, *explanatory* models (Q2) and *underlying* causal mechanisms beyond the phenomena and empirics. This especially became apparent in a key-issue in Greek philosophy: how to account for the related concepts of variation and manifestations of change; change of location (motion), growth and decay, change in quality and quantity. Variation and change had pejorative connotations and were indications of imperfection and unpredictability. Their existence was often denied, deemed impossible on metaphysical or logical grounds and reduced to non-change or otherwise corrected Starmans (2011a). Variation was considered to be a deviation from a rule or standard, which at best should be accounted for or explained for example by blaming unreliable sensory input. Against this background several appearances of causality occurred, that showed how gradually the worldview became more abstract and less tangible. The scientific worldview seems far removed from our daily experiences, intuitive concepts, common sense notions, and the natural categories we use to understand ourselves, our situatedness, and the contingencies of being. For example, Ionian and Dorian philosophers advocated a strong reductionism, distancing themselves from everyday perceptions and reducing the multiplicity of phenomena to first *causally active* principles or primary elements. Thales' solution (water) and that of Heraclitus (fire) still had some graphic "imagery", but Anaximander appealed to the abstract concept of "apeiron," (Q3) or the fundamental indeterminate. The Pythagoreans put the reality of numbers and numerical relations (Q1, Q4) above alleged material and observable objects. Eleatic philosopher Parmenides focussed on the material level. He became the most radical advocate of the immutability of being and denied the existence of variability and thus the primacy of the senses. Heraclites alternately defended the continuous flow of matter. Both had to come up with different explanations and different concepts of causality:

a formal, abstract account and a dynamic process approach. Clearly, the metaphysical foundation (Q1), abstract (Q3), largely deterministic (Q5) and qualitative character (Q4) were manifest already, usually at the token level (Q2).

Plato combined Pythagorean and Parmenidean ideas in his theory and showed little admiration for science that focused on the phenomena. His reliance on the metaphysical roots of causality was apparent and today Plato is especially recognized for declaring the metaphysical principle of causality, which roughly states that all that exists, becomes or changes will do so due to some cause, for nothing can exist, become or change without a cause. (Q1) However, it could be argued that Plato did a lot more than that such as deliberating the concepts of necessity and determinism (Q5) heralding the beginning of a theory of causality with a number of highly problematic fragments in his dialogue Phaedo. But his concept of causa formalis was highly connected to his theory of ideas, and therefore only slightly affected the philosophy of causality. All this is even more apparent in the work of his student Aristotle, who offered a much more systematic and influential treatment in the *Analytica Posteriora* and *Physics* with his doctrine of four causae ("aitiai"). As has been put forward in the Prologue to this essay a full explanation and understanding of a particular entity requires answering four key questions. This doctrine was at the heart of his metaphysics and epistemology, led to many interpretations and modifications in theology, philosophy and science until the 17th century, making Aristotle the founder of the theory of causality. Clearly it fits his qualitative physics (Q1, Q4). Intentionality / reasons were part of the concept (Q6) and also (Q7) was apparent due to the fact that Aristotle was the first philosopher of language and identified rhetorical devices, representations, shifts in the burden of proofs, when making causal inferences, depending on the type of dialogue involved for example in court, politics, ethics, biology. His pluralistic view and essentialism ("each to his own") permits or at least tolerates a more pluralistic view on causality. This is in full accordance with the traditional way of championing common sense thinking; he put the "essential forms" in phenomena (Q2, Q3), took great interest in the analysis of ordinary language (Q7), and at times showed a fundamentally empirical attitude. However, at the same time Aristotle struggled with the variation in the sublunary world and we found that variation and changeability of matter constituted an obstacle to formulating laws and undermined his axiomatic-deductive worldview. Determination was problematic and the concept of probability was not really developed yet. Aristotle's epagoge was not a genuine type of induction or generalisation, so with respect to (Q2) he was somewhere in the middle. However, we should go beyond Aristotle to do justice to the many early conceptualisations of causality, especially with respect to (Q4) and (Q2). Of course, the work of atomists and the Stoics should be mentioned here, who already explored the concepts of necessity, determinism (Q4) and exceptionless regularity (Q2). Indeed, the Stoics came up with a exceptionless regularity, however considered it primarily at the token level. The atomist movement claimed that reality consists of small particles (Q3) and the motions were determined and, in a way, necessary (Q5). Only the *clinamen*, the first shift of the atoms, allowed for some indeterminism (Q5), chance and possibility, without really advocating natural laws (Q2) a concept that would only arise in the 17th century. Indeed the concept of natural laws was strictly speaking absent in antiquity and came to the foreground in the seventeenth century, giving shape to the triptych "explanation, causality and laws", which has been continued even today as the notions of scientific explanations, natural laws, and causality are often studied as highly related or associated. No doubt, Aquinas strongly influenced thinking about causality and his famous five ways to prove the existence of God all contained different aspects of causality. What he actually did was further developing the causa efficiens and finalis. Important shifts took place during the scientific revolution where only causa efficiens remained, all teleology was banned. Aristotle's one powerful account was reduced to a truncated concept, that could easily be attacked or circumvented. For example, Francis Bacon, radically broke with this tradition claiming that science should discover the forms of the things, not Aristotelian essences, but immediate and concrete physical causes (Q1, Q3) based on empirical data, not intuitively postulated final causes; all teleology was claimed to be void here and should be limited to the explanation of human actors, their motivations and goals (Q6). His method of eliminative induction preceded 19th century work of John Herschel and John Stuart Mill and several of todays *difference making approaches* to causality. In other perspectives known since antiquity, causality has always been linked to determinism (Q4), physical and logical necessity (Q1). Descartes (geometric causality), Spinoza (logical causality) and Hobbes (materialist, atomistic account) were all similar in this respect. The fact that each conceptualization was closely linked to the metaphysical and epistemic views of the thinker in question is particularly apparent in the many theories of matter (corpuscular, atomic) that emerged in the 17th century. This applies even if one restricts oneself to the simple starting point that matter in a manner of speaking fills space and that matter is active in space. It may fill space continuously or discontinuously and may act dynamically with forces at a distance or kinetically with forces acting at contact, resulting in four different theories that could account for virtually all accounts of matter in the seventeenth century Snelders (2012). These include Descartes, Beeckman, Hooke, Boyle and the one of the statistician William Petty, who was one of the many scholars that created speculative atomist or corpuscular theories sometimes even with animated particles or just mathematical points. More importantly the associated concepts of causality were different as well, fully determined by the alleged structure of the cosmos, the processes and mechanisms. In an entirely different way, the connection between causality and theory is illustrated when dealing with the body and mind duality, that Descartes evoked. The question how two separate qualia like body and mind could be causally related is still relevant today in the philosophy of mind and in dealing with mental causation. Nicolas Malebranche famously came up with his occasionalism, denying a causal link (Q1, Q2) between both of them, but assuming that God would synchronize both (Q6), making God the causa efficiens rather than body, mind or both of them. Leibniz on the other hand agreed that no efficient causation between distinct substances was feasible. He did not to appeal to God, but postulated in his monadology that every single substance was endowed with the power (Q6) to produce changes in itself. Of course, Leibnitz tried to reconcile several conflicting ideas on matter, motion and associated cause and effect relations with one all-encompassing concept of causality, and his principles of sufficient reason and ideas about intelligibility of the universe are even today widely studied. As Hacking pointed out Hacking (1975) Leibnitz concerns with probability and uncertainty made him a precursor of statistics. In all these debates (Q5) was at stake, but increasingly (Q4). In fact, it was Galilei who actually replaced the "why" with the "how" question and brought great progress in kinematics and mechanics, but also changed the dominant qualitative approach to causality into a quantitative one. It was Newton, who could not come up with a mechanistic explanation and abandoned the idea of causality (Q1) taking into account only a mathematical description with no causal representation (Q7). Quantitative approaches (Q4) were introduced by Galilei and Newton, although at the same time causality as a metaphysical principle (Q1) became less apparent and needed. Unlike Laplace one hundred year later, Newton was not a full determinist (Q5), because in his celestial mechanics he needed divine intervention (Q6) to maintain stability in the cosmos. Of course, in this short section, we only mentioned the most dominant approaches and conceptual shifts to show the diversity in causality and for the purpose of illustrating our modest conceptual framework.

8. The mind, scepticism and elimination

In this section we will deal with the fact that in spite of the great role of causality in human behaviour and philosophy, it increasingly turned out to be a problematic notion in the course of the history of ideas and regularly hit a crisis. We will do so by briefly paying attention to the criticism of causality since the focus on the mind starting with the work of Hume, Kant and later Mach and Churchland. In the 18th century it was mainly David Hume (1711-1776) who sharpened the matter and challenged the (metaphysical) status of causality from his concept empiricism. According to Hume, causality does not exist in reality. There is only a constant sequence of phenomena, which in our minds are associated with habituation, then projected into reality, thereby explaining our "intuition" of necessity. David Hume famously challenged the (metaphysical) status of causality based on his concept empiricism. According to Hume causality does not exist in reality, there is only a constant regular sequence of phenomena. In our thoughts, due to natural habit they are associated with physical necessity in our minds, and projected on reality, thus explaining our "intuition" of necessity. Induction and causality were thus inherently connected and both not rationally justifiable. After Hume's scepticism, it was mainly Immanuel Kant (1724-1804) in the 18th century who tried to restore the notion and pursued a reconciliation between metaphysics and natural science. He famously acknowledged that Hume had pulled off his dogmatic slumber by considering causality as synthetic a priori knowledge and calling it a fundamental category of human knowledge, which is a necessary condition for observation, experience and scientific knowledge. So due to his top-down approach he could declare void Hume's problem of induction, which obviously is a bottom-up issue and also guarantees the eternal validity of Newtons mechanics, that seemed to have lost its philosophical foundations since Hume's sceptic approach. Of course, to achieve this and to reconciliate older conceptions of causality Kant had to postulate a dualism of the real noumenal world and the world of phenomena.

Be that as it may, with the anti-Kantian scientific conception of logical positivists in the 20th century, this conception of causality also proved problematic, not least because of the successes of quantum mechanics, in which causality was no longer regarded as a Kantian building block of reality. In the 19th century it was then, among others, the physicist / philosopher Ernst Mach (1838-1916) and the statistician / philosopher Karl Pearson (1857-1936) who manifested themselves as prominent anti-causalists. However, the most important criticaster presented himself in the 20th century. It was Bertrand Russell (1872-1970) who, in 1913, published his article "On the notion of Cause" in The Proceedings of the Aristotelian Society and stated that "[t]he law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm." Although Russell took a more mitigated point of view in his later work (Russell, 1943) from the point of view of natural science there was some sense in it and there was great progress in many sciences including physics, chemistry and engineering without any feeling the need of explicit causal models.

It is important that this attitude towards causality does not stand on its own. Over the years, many concepts and notions related philosophical themes have been discredited or, at best, given a specific abstract or mathematical interpretation. This applies, among other things, to notions such as space, time, movement, mass, but especially to concepts such as meaning, intentionality, spirit, free will, consciousness and personal identity. According to some scientists and naturalist philosophers, all philosophical problems will ultimately be unravelled and revealed by science. If the problem is well defined, it will be analysed and then resolved. If it is not well defined, then it is dismissed as a pseudo problem or as meaningless. During that process, philosophical reflection can at most lead to a kind of pre-scientific theories, which may have some explanatory power or practical utility, but which will ultimately be replaced by true scientific knowledge Starmans (2011a). The concepts that play a role in this philosophical reflection will then usually have to leave the field. This tendency to purify science and its language from metaphysical concepts, common sense notions, natural categories and everyday experiences has a notorious high point in the views of the philosopher and neuroscientist Paul Churchland (1942), who wants to radically deal with a tradition that sometimes becomes pejorative referred to as "folk psychology". People try to understand, explain and predict the behaviour of themselves and others in terms of causally relevant factors, such as motives, intentions, beliefs and obligations. Churchland argues for a radical "eliminative materialism" regarding these propositional attitudes and argues that "folk psychology" including the notion of consciousness, is completely wrong with the human mind and its internal processes. He also regrets the preoccupations of philosophers with language and their supposed crucial significance for thinking. Developments in neuroscience will, according to Churchland, lead to the elimination of these "errors", which he considers to be just as relevant to science as Stahl's 18th-century phlogiston theory of modern chemistry, or medieval views on witchcraft to contemporary psychology Starmans (2018f).

The extreme vision of Churchland fits into a long-term development in the history of ideas that started with the pre-Socratics and reached a peak in contemporary naturalistic / physicalist epistemology. A consequence of this is that the current worldview has lost much of its portrayal. On the one hand there is the everyday, familiar world of phenomena, with its experiences (perceptions, impressions), representations and ideas and with its (postulated) concrete objects. On the other hand, there is the scientific worldview with its abstract, often mathematical models, representations of "real" reality, which is supposed to hide behind these experiences, and which is supposed to cause or explain them. Not only do the experiences, the phenomena that occur to us, do not seem to be a reliable basis on which to base scientific theories, the intuitive concepts and natural categories also seem to have little in common with the underlying mechanisms, abstract principles and laws that the "real" world as described by the language and nomenclature of science. It is a next phase in creation distinctions and schisms that originated in the 19th century as explained in the Prologue. It is also clear that this vision cannot be reconciled with the linguistic perspective discussed above. The evolution of causality must also be understood in this field of tension. After the mentalistic interpretation of Hume and Kant, the sledgehammer attacks of Mach, Pearson and Russell, and the Churchlandian urge to eliminate obscure concepts, the fate of the concept of causality in science seemed sealed.

9. Variation, uncertainty and statistics

In the Prologue we described how romantic conceptions of science (the Great Chain of Being, “was die Welt am innersten zusammenhält”, the causes of things) arose on the eve of famous 19th century philosophical debates on knowledge and methodology. These debates could only emerge against the background of the rise of modern science in the same era: the proliferation of new disciplines, fragmentation of knowledge, spectacular progress in mathematics and physics, recurrent foundational crisis in recently emerged disciplines like psychology, sociology and economics and a process of historicization of the worldview Starmans (2011a). This would spark the probabilistic revolution, which in turn induced real progress in the philosophical concept of causality and reinforced its entrance into the realms of science, due to the convergence of various developments. First of all, according to the philosopher Ian Hacking, “an erosion of determinism” took place Hacking (1989). Acquired knowledge about the capriciousness of (living) nature and the multiplicity and variation of its manifestations, coupled with a historicization of the world view, formed an insurmountable obstacle to a deterministic view of reality. Variation, change and uncertainty became central concepts and -partly due to the cross-pollination between biology and statistics- a probabilistic worldview emerged. With that, the final blow seemed to have been inflicted on a conception of causality as physical or logical necessity or exceptionless regularity that was usually associated with a deterministic worldview. In fact, an emancipation process in thinking about uncertainty took several steps. Adolphe Quetelet showed how variation and uncertainty in society could be tamed and caught in laws, using very elementary statistics and canonizing the Gaussian distribution. Then Francis Galton and especially Karl Pearson would give uncertainty an important place in the scientific worldview and succeeded to encode, variation and change with probability distributions, mainly inspired from biological research. After that, Fisher would for a substantial part develop statistical inference and experimental design, including randomisation, interventions, et cetera. Due to Maxwell, Boltzmann, Gibbs and others statistical concepts would enter physics (kinetic theory of gases, statistical mechanics). Finally, physicist Niels Bohr would argue with his Copenhagen interpretation of quantum mechanics that uncertainty is a building block of nature, irreducible and cannot be traced to a lack of knowledge Starmans (2018f). The erosion of determinism asked for *a new language of science* that could deal with all these aspects and this language was not available yet.

This brings us to a second development, that arose from a methodological point of view. As pointed out in Section 5 virtually all research questions are explicitly or implicitly causal. However, in the newly established “variation and uncertainty rich” empirical sciences, like biology, genetics, agricultural science, psychology, sociology and economics there was a need for a *new, constructive notion of causality*, which had separated itself from both old metaphysics and Laplacian determinism. Philosophical attempts to anchor causality in a more pragmatic and experimental context can be found in the work of, among others, and especially John Stuart Mill (1806-1873) and C.S. Peirce (1839-1914). The former considered causality from the totality of circumstances that had to be known, checked or manipulated in order to establish causal relationships or an intended causal effect. His *System of Logic* (1843) contains the famous “five methods of Mill” and was in fact a methodological handbook avant la lettre, in which the author tried to bridge the gap between abstract epistemology and actual scientific /experimental practice, especially with regard to thinking about causality. The influence of empiricism and positivism was unmistakable. All this turned out to be even more relevant, because many new and independent empirical sciences had little status, no established methodology or foundations, and therefore they all experienced understandable and necessary foundational crises. The process of disentangling the speculative philosophical tradition required an operationalized concept of causality. Object, purpose and method had to be determined and a language was sought that could do justice to variation and covariation, uncertainty and the inherently causal associated research questions; a formal method that allowed for a causal interpretation that matches the specific causal questions within that scientific field. They would find this status in the newly established field of statistics pioneered by Quetelet, Galton, Pearson, Yule, Edgeworth and later Fisher, that provided them with the required language and methodology. This is to be considered a third development that characterised causality in the 19th century.

Finally, it could be said that statistics really played a significant role in attempting to defuse the crises of causality we outlined in Section 4. Of course, it could be argued with some good will that the link between statistics and causality was already evident in the 17th and 18th centuries, such as with Pascal's famous wager, Bernoulli's law of large numbers, but especially with John Arbuthnot's causal interpretation of the fixed proportions of boys and girls in birth rates and of course in Thomas Bayes' exploration of the reversal of probabilities. However, the specific combination of factors in the 19th century, was needed to really enable statistics to enter into a sort of progressing liaison with causality, because the many statistical techniques were nearly without exception initiated, developed and validated in close interaction with everyday practice, the specific causal research questions arising in particular new disciplines. However, there was no one-way traffic and in fact, there was an unprecedented cross-fertilisation between statistics and the new "variation and uncertainty rich" sciences as of the very start of these sciences Starmans (2018c). Because of this interaction, these sciences were a substantial part of the probabilistic revolution. As of then, and up to this very day, statistical techniques are typically developed, assessed, revised or rejected depending on their "causal fitness" or "causal suitability". The innovation was pushed forward by the fact that statisticians proved to be increasingly competent in searching, creating and analysing new sources of variation and covariation, understanding the importance of intervention and by creating more complex methods and techniques that did justice to this. Of course, developing mathematical techniques to safeguard valid inference completes this list. It cannot be overemphasized that all this was mainly due to the fact that research was not based on armchair philosophy or toy examples, that immediately ran into trouble once scaled up or extended to a real-life problem, but because real-life causal practical problems presented themselves that could not be adequately addressed with existing methods Starmans (2018f,c). We have to restrict ourselves to a few examples here. It all started with the anti-causalist par excellence Karl Pearson. Immediately after Pearson's classic papers, a famous polemic started with Udny Yule on the interpretation of correlation. Then the partial and semi-partial correlation came through, criticized on causal grounds by Burks (1928). A next step was regression analysis and more importantly Sewall Wright's method of path coefficients, which led to important causal debates Starmans (2018d); the resulting approach of structural equation models (SEM) is virtually synonymous to causal modelling according to many researchers in social sciences and economics. In the 1920s, the analysis of variance came to fore, based on Fisher's ideas on relating causality to specific aspects of experimental design such as intervention, randomisation, blocking, published in his famous book *Statistical Methods for Research Workers* (1925). It would pave the way for a rich causal tradition in the methodology of social sciences resulting in many studies of research design by Blalock, Cook, Campbell and others, who all combined the philosophical work of Stuart Mill with statistical techniques developed by Fisher and others.

But also factor analysis and principal component analysis were considered as techniques which could deal with hidden, abstract variables, that were not directly measured and referred to abstract or postulated entities (Q3), but which were still considered to be causally effective. Of course, we should mention the famous "social facts" that Emile Durkheim postulated as causal entities which are irreducible to individual entities. The most historically salient example, however, relates to the fact that the aforementioned new sciences used data that were interpreted as hierarchical; they are nested, embedded, or layered, i.e. have different levels. The interdependencies and interactions between the different "levels" are essential in causal research questions and theories. The aim to involve the various "levels" simultaneously in an analysis led to many forms of multilevel analysis. Whether it concerns multi-level models in the strict sense, hierarchical (linear) models, nested data, mixed models, classical split-plot designs, random coefficient or random-effect models, repeated measures, et cetera. Searching for and creating new sources of (co-)variation leads to models with more random effects, fewer fixed effects and a more sophisticated analysis of residual error, all within and between the different levels. Finally, it should be noted that many of these and other techniques can be linked to classical paradoxes and anomalies, related to confounding, spurious correlation, Simpson paradox Chambaz et al. (2020), Robinson's ecological fallacy Robinson (1950), et cetera. Today, all these methods dominate experimental and observational studies and form the basis for causal statements in many empirical disciplines. Without this liaison with statistics, the concept of causality would have long disappeared from the sciences in the light of the identified crises briefly sketched in this essay. Causality became quantitative (Q4), probabilistic (Q5), allowed both for token and type inference (Q2), for concrete and abstract entities (Q3) and was strongly associated with knowledge, the mind, and our understanding of the world rather than with physics itself (Q1,Q6), for example in Bayesianism and represented with / in the language of probability (Q7). With respect to (Q7) it should be noticed that for many years statisticians and methodologists have tried to represent, encode and grasp essential aspects of causality with the aforementioned techniques or more generally put, with probability measures, conditional independencies, smart factorial designs, analysis of contingency tables, formalizing (assumptions of) the data generating process, randomisation, several validity and reliability concepts, estimation techniques, or more generally put statistical inference. So, the fact that today nearly all disciplines have experienced a probabilistic turn, could arguably support the conclusion that they have also taken a causal turn. Still it should be noticed that in theoretical statistics for many years causal talk was missing, i.e. the aforementioned techniques were used without directly referring to the word causality and without directly formalising it. Only the last three decades theoretical statisticians have started making explicit references to causal inference (Henin, 2019), van der Laan and Rose (2011, 2018). They typically did so without representations that exceeded the language of probability (Q7). For many applications in the sciences this was efficacious, and many "anticausalist" positions we outlined in Section 4 and in Section 8 emphasizing its alleged obsolete character, basically concerned primarily (Q7). However, the ambitions of AI would move the question of causality to a next level, especially regarding (Q7).

10. Causality, AI and data science

In 1985 the Australian physicist and AI pioneer Peter Cheeseman published his famous, but somewhat polemic article *In Defense of Probability* in the Proceedings of the Ninth International Joint Conference on AI Cheeseman (1985). Today, the IJCAI is still the most authoritative mondial AI conference. In his contribution the author argues, among other things, that probability theory and probabilistic methods are sufficient to achieve automatic reasoning with incomplete and uncertain knowledge and the common sense reasoning envisaged in the AI. He criticizes the at that moment dominant logical tradition within symbolic AI and states that all criticism on probabilistic approaches stems from misunderstanding and ignorance. The "sources of error" are then analysed in a rigorous manner; confusion about a frequentist concept of chance versus "measures of belief", confusion about absolute and relative probability, confusion about probability and the uncertainty of that probability. He also attacked the critics about their - in his eyes - obvious misunderstanding of the Bayesian foundations. Above all, Cheeseman regrets the "proliferation of representation languages's with associated inference procedures", all extensions of classical logic, which are unsuitable and unnecessary for the realization of the ultimate Project of the AI. "Probability is all that's needed," according to the author.

In a way this is a remarkable contention. For instance, at that particular moment the graphically oriented probabilistic (Bayesian belief) networks didn't exist yet, they would take shape no earlier than the late 1980s with the work of Richard Neapolitan, David Spiegelhalter and Judea Pearl. Moreover, in the mid-1980s AI's was still firmly in the hands of the classical symbolic knowledge representation, the declarative, logical programming language PROLOG was regarded as the "lingua franca" of AI's and the Fifth Generation Project of Japan represented the high ambitions in this respect. All this led to much research and progress in logic. Numerous modal logics (epistemic, deontic, temporal) were developed, non-monotonic logics made their appearance, which remained influential well into the 90s in spite of Cheeseman's paper. What's more, the symbolic logic, language and semantically oriented AI was also much more "salonfähig" among cognitive psychologists, philosophers of mind and linguists who embraced the ideals of Strong AI. A theory of the human mind was central, "high level cognitive functions" had to be represented, symbol manipulation was a sufficient and necessary condition for intelligence and reasoning. A rich logical language should provide the foundations and neural networks were out of the question and, more generally, subsymbolic AI, which - rightly or not - also included the probabilistic methods, played a second violin. Modelling *causality as an indispensable quality* of the human mind was not high on the research agenda, neither in the logical nor probabilistic tradition. Also, in the Philosophy of AI Haugeland (1985), J. (1993) there was little interest in the rich literature on causality, as outlined in this thesis. All this may be a meaningful observation but will not be discussed here any further. Still it is remarkable, especially for those committed to the ideals of the project of Strong AI, which is not likely to succeed without a proper account of causality.

Still *In Defense of Probability* can be called visionary in various respects at the same time. The aforementioned probabilistic networks would soon come to the fore allowing for advanced reasoning with uncertainty. More importantly they allowed for tacitly or implicitly making *causal inferences* or *answering causal research questions*. Causality was for a substantial part encoded, represented in the language of probability (Q7). Furthermore, as of the ninetieth subsymbolic AI started booming and the neural net winter of the seventies and eighties seemed definitely over. In a sense, it could be claimed that the rise and successes of Deep Learning are the crown on this work. Moreover, precisely these results dominate current AI-data science debate regarding the risks and opportunities of AI and the call for Responsible and Explainable AI. Paradoxically, the widespread concern regarding many ethical objections to the opaque and inconceivable, seemingly objective and value-free deep learning algorithms, which influence the lives of many without human intervention, illustrates the success of subsymbolic AI. We are not dealing with "drawing table" work, armchair philosophy or "toy examples", but working systems that can radically change society. Furthermore, the fact that statistical learning, machine learning, computational intelligence and data mining, which paved the way for current data science, are all probabilistic emphasized Cheeseman's visionary position. The same applies to Shannon's information theory as a basis for many learning algorithms. Finally, it must be noticed that although Cheeseman was of course not the first to argue for Bayesianism, he empathically stated that it is a "major aim" of the paper "to put forward the older view", namely the work of Bayes and Laplace. Only in the following years Bayesianism would break through, within mathematical statistics, within computer science / AI, but also in the philosophy of science / knowledge theory (confirmation theory), albeit in entirely different ways.

Of course, Cheeseman's position did not remain undisputed and some objections that are relevant for the theme of causality will be discussed briefly here. First of all, the history of the AI's has shown that his claim lacks nuance. Postulating a probability statement as a necessary and sufficient condition for reasoning in AI involves too rigid a position. The symbolic AI has also made great progress in the decades that followed and today the dust clouds seem to have moved somewhat. More symbiosis can be detected, which is also visible at large AI conferences. A complete synthesis is perhaps utopian and in real-life systems that really matter, it is perhaps better to postulate a logic-probability tradeoff in which maximizing one at the expense of the other inevitably leads to a less functioning system. Historically, this all fits better with the old ideal of a "calculus ratiocinator" from Leibnitz in the 17th century and especially with that of George Boole, who explicitly linked "the laws of thought" to a language in which both logic and probability are taken into account. Recent advances in Hybrid AI are rooted in this tradition.

There are, however, other considerations. Ironically, two "sources" are already in the very same paper of Cheeseman. Among other things the author refers to an early study by Judea Pearl on causal reasoning from 1983 and also to the groundbreaking psychological research by Tversky and Kahneman, as described in *Judgement under Uncertainty: Heuristics and Biases* from 1974 Tversky and Kahneman (1974). These and numerous other highly relevant psychological studies show that man is not an intuitive statistician, whether in causal reasoning or otherwise. He often makes unquestionable mistakes against the laws of probability, even after years of scientific training, and those mistakes and "biases" should of course be avoided or corrected. Sometimes, however, man does not follow probability theory at all, but uses heuristics and analogies, which prove to be very suitable for many intelligent everyday tasks, which are (evolutionarily speaking) highly successful, a efficacious quality of the human mind, and therefore relevant for AI. It brought Kahneman the Nobel Prize in Economics in 2002. For some philosophers, all this would lead to a renunciation of probability theory and its supposed counterintuitive character as cornerstone of any theory of reasoning and inference. A more sober conclusion is that people simply are not natural "probability calculators" and that all related skills should therefore be learned and trained.

A-fortiori all this has serious consequences for conceptions that directly concern the moral experience or have a strong moral dimension; responsibility, justice, reasonableness, reliability, trust, power, democracy, care, safety and risk. In a rationalized society these concepts are intrinsically probabilistic and appear to be less and less consistent with the familiar categories of thinking and acting, the individual moral experience, the tried and tested imperatives for ethics and the associated moral support. Indeed, in the light of Kahneman's insights and the current call for Responsible and Explainable AI, a great deal of tension is apparent.

The second "source" that is already in Cheeseman's paper concerns Judea Pearl. Although research on causality continued to play a role in the symbolic AI, including logical approaches in Knowledge Representation and including commonsense representations in Qualitative Reasoning, today the issue of causality in AI is highly dominated by probabilistic approaches and more particular by Pearl. As of his study on Causality in 2000 he criticizes all existing approaches including the Bayesian Networks, pioneered by himself in the nineteen-eighties and nineties, especially regarding (Q7). In this study and more recently in Pearl and MacKenzie (2018) he strongly decries research in statistics over the last hundred years for neglecting, undervaluing and misrepresenting causality, which he even considers a major threat for progress in science. This is remarkable because his own approach is obviously probabilistic as well, but here we will restrict ourselves to another more relevant aspect of this recent work. In the final chapter of Pearl and MacKenzie (2018) he explains in detail that the project of AI in general and the ambitions of Strong AI in particular need a specific causal approach and he criticizes the successes of Deep Learning and the associated problems as deviations of the ideals of real AI. Pearl's view that statistics "contains" too little causality should be viewed against the background of its higher goal: saving the high ambitions of strong AI. More than in estimation theory or inferential statistics, he seems interested in equipping robots with a human notion of causality. In contrast with the view of Churchland, as outlined in Section 8, Pearl considers language crucial for thinking and an essential condition for moral intelligent agents. It would appear that the present debate partly fits in with the long-standing AI controversy to represent or not, which has been sharply articulated in the classic publication by Rodney Brooks Brooks (1991). Should we try to formalize causality in the object language (for example by an operator) or in the meta-language by research design, mechanisms of data generation, by more advanced statistical techniques or by recognizing that it concerns specific contexts and procedures that can be represented in specified language games or dialogue games. Regardless of the "represent or not" contrast, you can be a causalist without explicitly representing it, you can even be a causalist without wanting to use the concept, as was the case with Karl Pearson. In an even more recent publication Pearl (2019) the author argues that to successfully solve this, a dialogue between man and computer using human causal language is required. He outlines how a three-layer causal hierarchy (association, intervention an counterfactual) and several tools, including do-calculus, algorithmizing of counterfactuals, and causal discovery are corresponding with "seven cognitive tasks". According to the author these are necessary steps in realizing the ambitions of Strong AI. As this short paper lacks the many unpleasant attacks on statistics and its contributions to the history of science presented in Pearl and MacKenzie (2018), but synthesizes and outlines the authors ideas on causality of the past 25 years, it could or should well be a hinge point and calibration point in the contemporary AI-data science debate.

11. Epilogue

In this final section we wrap up a little and make some comments on dealing with causality in the AI-data science debate, its problems and challenges, based on our historical-philosophical view.

Firstly, it should be noted that causality is still en vogue, at least in statistics and AI, despite the sketched crises and disunity. Still there is little cross-fertilization or cooperation between the different (formal) approaches to causality and the burgeoning recent literature on causality has only modestly influenced actual research practice -and more generally- research methodology. What is left is a plethora of sophisticated approaches (not mutually exclusive or totally exhaustive) that all try to grasp the "essence" of causality. The proclamation of pluralism as a last resort or escape route becomes almost inevitable, and any pursuit of naive unification or unity seems illusory. For those who want to solve the problems of the AI-data science debate by relying on existing literature on causality this is not a convenient start.

Secondly, it is quite obvious that formal probabilistic approaches in causality, such as Pearl and MacKenzie (2018); Pearl (2019); van der Laan and Rose (2018); A. and Robins (2019) made spectacular progress, the last three decades. Still in view of the AI-data science debate this is only part of the story. In many applications causality occurs at the token-level (Q2), it is qualitative (Q4) and deterministic (Q5), adequately represented in everyday natural language (Q7), using implicit or explicit causal language as a primitive or commonsense notion. Neither postulating physical mechanisms (Q1), nor dealing with "reasons as causes" (Q6) seems problematic. The same applies to making and understanding causal claims on abstract entities, multi-dimensional concepts or concepts that are not easily measurable or even "ill-defined" from a scientific point of view as such, the concept has many professional applications in the legal, medical or technical domain. And it is used by citizens, voters, judges and attorneys to *understand and explain the behavior of artificial systems*, asses the fairness of the algorithms and judge their moral acceptability. Natural language is not intrinsically vague, unreliable and error-prone; it is not only the starting point of the analysis, it could well be the level at which we should conduct our analysis. To grasp causality, to meaningfully use and understand it is to be engaged in a language game, a dialogue where interlocutors are trying to resolve a dispute, to make a decision or negotiate. According to the rules and conventions of that specific language game they make claims, arguments and counterarguments, shifting the burden of proof, making commitments and change the world by performing speech acts. Of course, this is only one way to look at it, as pointed out in Section 5 and Section 6, but in the current discourse on causality, this is sometimes slightly underexposed.

Thirdly, one should acknowledge that, as pointed out in Hacking (1989) and Starmans (2018e) statistics has always had a rather problematic relation with ethics for several reasons. Some ethical problems of statistics are intrinsic and not easily to be circumvented. Since the probabilistic turn, which has now taken place in almost all sciences, not only the methods used, but many concepts and concepts can only be interpreted meaningfully or with the help of probability theory, statistics, probability distributions or parameters of data-generating functions, parametric or semi parametric models, estimation procedures etc. As described in Starmans (2011a), this was accompanied by a further decline in the portrayal of the world view and the familiar categories of thinking and acting. The rationalization of society and associated institutions, models of policy, administration and organization that are based on these principles in order to get a grip on uncertainty, have "a fortiori" consequences for conceptions that directly concern the moral experience or have a strong moral dimension: responsibility, justice, reasonableness, reliability, trust, power, democracy, care, safety and risk. These often proved to be less and less in accordance with (the familiar categories of) the individual moral experience, the tried and tested imperatives for moral action and the associated moral support. But other factors such as habituation, habit formation, adaptation, coercion, profit, utility do also play a role and are now undergoing a transformation due to the rise of AI-data science debate, where machines are supposed to be equipped with that, especially if one bears in mind the specific nature of the technological knowledge domain, including associated values. And again, stakeholders in this particular technology debate use (the familiar categories of) the individual moral experience and values to understand and explain the behavior of artificial systems, asses their fairness and judge their moral acceptability. As a result, formal approaches to causality, which are based on probability and statistics, do inherit their intrinsic properties and problems with ethics.

Fourthly, it seems that high expectations on transparent and explainable AI should be tempered in view of a key distinction made by the aforementioned physicist and philosopher of science Hans Reichenbach, the logic or context of discovery versus the logic or context of justification. Being a logical-positivist he wanted to give a rational reconstruction of science, which means that philosophers should primarily deal with the process of justification, rather than trying to conjecture how scientists actually came to their results or trying to find an algorithm for scientific discovery like the empiricist Bacon and the rationalist famously tried 300 years earlier. We will not go into the discussion whether of not this distinction would or should be made today, but many agree that it was necessary or at least beneficiary at that particular point in history. This distinction can be extended or generalized and applied to the current AI-data science debate and the experienced problems regarding autonomous systems and the call for transparency and explainability. The algorithms are complex, opaque, difficult to understand, there are no internal representations that "make sense" and it is unclear which data is used, how it is used, which mechanism leads to certain conclusions and for which purpose it is optimized. Internal processes are difficult to trace and track, to represent or express. A causal explanation, based on an operating technology tracking the mechanism or intentionality, is problematic, however a justification is much required for all kind of reasons: economic, scientific, legal and ethical. What we are actually doing for the main part is just giving a rational construction, that may convince the user or client, that is in harmony with regulations and if necessary, that can be defended in court, arguing that we did not trespass, have been considerate and properly conducted any language game, dealing with specific criteria of fairness, reasonableness, rather than understanding the real causes of things. The question whether this Reichenbachian gap can / should be bridged may count as another challenge to the field of AI and data science.

A fifth remark combines the third and the fourth. We are increasingly facing a contemporary Euthyphro dilemma. In Plato's famous eponymous dialogue, Euthyphro was on his way to justice to report his father, who he deemed guilty of having killed one of his employers due to his negligence. Euthyphro did so appealing to the Good and the will of the Gods. Socrates – like he usually did - forced his conversation partner into a corner and undermined his source of knowledge and moral foundation. How could Euthyphro know the will of the Gods with regard to the Good, and above all: is something good because the Gods want it, or do the Gods want it because it is good? Much has been written about this by theologians and philosophers in recent centuries.

In addition, someone who used to consult the Delphi oracle to solve the first part of the problem, because he also had to overcome an interpretation problem, since the message was generally very dark and vague and required high priests to decipher it. Anyone who wants to question the source of knowledge and the moral foundation in the AI's data-science debate, asks an analogous question: Does the autonomous and inscrutable system wants, chooses, decides or recommends something because it is "good", or is something "good" because the autonomous and inscrutable system wants, chooses, decides or orders this? That question is no longer absurd, especially since the principle of intelligibility that has been at the basis of philosophy and science since Parmenides is at stake here. In this way, each system, whether based on Deep Learning or not is to be considered also oracle language. More and more AI-researchers realize that the inability to answer these causal questions may affect the success of the entire AI-project, especially in view of the stakeholders that play their role in any technology debate. Also, for ethics this is a big challenge and new playground, because increasingly a purely fixed communis opinio conception of ethics ("we don't want that") is inadequate and sometimes even obsolete. A naive ethic that wrongly suggests an unproblematic moral foundation, building up narratives with heavy ethical appeals like in O' Neill (2016), will rather be a prelude to ideologically driven and politically motivated debates or cultural pessimism than a contribution to an untroubled discourse. A modest plea for a meta-ethical perspective Starmans (2018a) could be an alternative.

Finally, we return to Prometheus. In the Prologue we stated that in the era of AI and data science a new chapter is being added to the genealogy of the ancient myth. Or rather, the process of retelling, re-creating and reinterpreting the myth of the unleashed Prometheus has entered a new phase. Prometheus' first project was complicated and thorny and had already caused many problems, but now that he has definitively thrown off his chains, he is ready for a second project and many are reluctant to embrace this. Building autonomous robots with a "mind" that possesses human qualities such as consciousness, emotions, language and morality. Building machines that produce knowledge beyond our comprehension. Building systems based on omnipresent data and intelligent, opaque, "black box" or deep learning algorithms; systems which may determine, monitor, assess, convict man, work in synergy with him, but may also dominate or replace man, who deliberately or not may have created his evolutionary successors. One need not embrace speculations on the singularity to acknowledge that the principle of intelligibility is at stake and epistemology may well face a crisis too. There will be no insight into the true nature of the phenomena, the Great chain of Being, "was die Welt am innersten zusammen halt", and now we may even no longer know the "causes of things", even in the fragmentary domains with their pieces of knowledge, that we could understand so far. In 1936 on the eve of the Second World War, it was the German Philosopher Edmund Husserl who wrote *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie: Eine Einleitung in die phänomenologische Philosophie* ("the crisis of european sciences and transcendental philosophy; an introduction to phenomenological philosophy). Among other things he criticised the fragmentation of the individual professional disciplines, which had rapidly emerged as "unselbständige Zweige der Einen Philosophie" ("immature branches of the One Philosophy") without unity and foundation, but also the deterioration of the image of the world view, the associated gap between daily experience and (the language of) science, that we described in Section 8, the decoupling of human experience, knowledge and the meaning of existence. It was not just Husserl's nostalgia for a bygone era of Romantic science and Shelley's epistemology. He was actually looking for a new concept of rationality, but all in all, his writings were an ultimate attempt to restore the old position of philosophy. In fact, the crisis he detected was a direct consequence of the methodological crises we outlined in the Prologue. The problems and challenges regarding causality, statistics and the foundational crisis of AI and data science, outlined in this essay, should be considered in this perspective.

References

A., H. M. and Robins, J. M. (2019). *Causal Inference*. Chapman & Hall/CRC, Boca Raton.

Abrams, M. H. (1953). *The Mirror and the Lamp. Romantic Theory and the Critical Tradition*. Oxford University Press.

Anderson, C. (2008). The end of theory: the data deluge makes the scientific method obsolete. *Wired Magazine*, 6.

Anjum, R. L. and Mumford, S. (2018). *Causation in Science and the Methods of Scientific Discovery*. Oxford University Press.

Berlin, I. and Hardy, H. (1999). *The roots of Romanticism*. Princeton University Press.

Bostrom, N. (2005). A history of transhumanist thought. *Journal of Evolution and Technology*, 14(1).

Bostrom, N. (2014). *Superintelligence: paths, dangers and strategies*. Oxford University Press, USA.

Brockman, J. (2015). *What to think about machines that think. Todays leading thinkers on the age of Machine Intelligence*. Harper & Collins, New York.

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence*, pages 1–3.

Brynjolfsson, E. and McAfee, A. (Norton). *The Second Machine Age: work, progress and prosperity in a time of brilliant technologies*. 2015.

Burgess, S. and Thompson, S. G. (2015). *Mendelian Randomization, Methods for Using Genetic Variants in Causal Estimation*. Taylor & Amp.

Burks, B. (1928). On the inadequacy of the partial and multiple correlation technique. *Journal of Educational Psychology*, pages 532–540.

Burwick, F. (1982). The language of causality in prometheus unbound. *The Keats and Shelley Journal*.

Chambaz, A., Drouet, I., and Memetea, S. (2020). Simpson's paradox, a tale of causality. *Journal de la Société Française de Statistique*. Special issue Causalité.

Chambaz, A., Drouet, I., and Thalabard, J.-C. (2014). Causality, a trialogue. *Journal of Causal Inference*, 2(2):201–241.

Cheeseman, P. (1985). In defense of probability. In *Proceedings of the Ninth International Joint Conference on AI (IJCAI, 1983)*.

Clark, A. (2003). *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford University Press.

Domingus, P. (2016). *The Master Algorithm How the Quest for the Ultimate Learning Machine Will Remake Our World*. Basic Books.

Franklin, S. (2014). *The Cambridge handbook of Artificial Intelligence*, chapter History, motivations, and core themes. Cambridge University Press.

Glymour, C. (2001). *The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology*. MIT Press.

Glymour, C. (2019). Review of: Rani Lill Anjum and Stephen Mumford, Causation in science and the methods of scientific discovery, Oxford University Press, 2018. *Notre Dame Philosophical Reviews, an Electronic Journal*.

Goethe, J. W. v. (1808). *Faust. Eine Tragödie*. Reclam Verlag, Berlin. Kapitel 4.

Goethe, J. W. v. (1982). *Theory of Colors (1810)*. MIT Press.

Habermas, J. (1981). *Theory des Kommunikativen Handelns*. Frankfurt Am Main: Suhrkamp Verlag.

Hacking, I. (1975). *The emergence of probability*. Oxford University Press.

Hacking, I. (1989). *The Taming of Chance*. Oxford University Press.

Halevy, A., Norvig, P., and Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(3):8–12.

Haraway, D. (1985). A cyborg manifesto: Science, technology, and socialist-feminism in the late twentieth century. *Social Review*, 80:65–108.

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. MIT Press.

Hayes, P. (1977). *The naïve physics manifesto*. University of Essex.

Hayes, P. (1978). In defense of logic. In *Proceedings of the fifth international joint conference on Artificial Intelligence (IJCAI)*, volume 1. Cambridge University Press.

Heidegger, M. (1954). *Die Frage nach der Technik*. Grin Verlag.

Horkheimer, M. and Adorno, T. W. (1972). *Dialectic of Enlightenment*. New York: Herder and Herder. Translated by John Cumming.

Illari, P. and Russo, F. (2016). *Causality: Philosophical Theory meets Scientific Practice*. Oxford University Press.

J., C. (1993). *Artificial Intelligence: A Philosophical Introduction*. Blackwell.

Jacobs, J. D. (2017). *Causal Powers*. Oxford University Press.

Kern, S. (2014). *A Cultural History of Causality: Science, Murder Novels and Systems of Thought*. Princeton University Press.

Krüger, L., Daston, L., and Heidelberger, M., editors (1981). *The Probabilistic Revolution, Volume I: Ideas in History*. MIT Press.

Krüger, L., Daston, L., Heidelberger, M., Gigerenzer, G., and Morgan, M. S., editors (1987). *The Probabilistic Revolution, Two Volumes*. MIT Press.

Lovejoy, A. O. (1938). *The Great Chain of Being: the study of the history of an idea*. Harvard University Press.

Macagno, F., Walton, D., and Reed, C. (2017). Argumentation schemes. History, classifications, and computational applications. *Journal of Logics and their Applications*, 4(8):2493–2556.

Mackie, J. L. (1980). *The Cement of the Universe: a study of causation*. Oxford University Press.

Mayer-Schönberger, V. and Cukier, K. (2013). *Big Data: A Revolution that will Transform how we Live, Work and Think*. Houghton Mifflin Harcourt.

McCulloch, W. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5(115–133).

Mumford, S. and Anjum, R. L. (2011). *Getting causes from powers*. Oxford University Press.

Mumford, S. and Anjum, R. L. (2013). *Causality, a very short introduction*. Oxford University Press.

O' Neill, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group.

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge University Press.

Pearl, J. (2019). The seven tools of causal inference, with reflections on machine learning. *Communications of the Association for Computing Machinery*, 62(3):54–60.

Pearl, J. and MacKenzie, D. (2018). *The book of Why: the new science of cause and effect*. Basic Books, New York.

Pearson, K. (2004). *The Grammar of Science (1892)*. Dover Publications.

Plessner, H. (1928). *Die Stufen des Organischen und der Mensch: Einleitung in die philosophische Anthropologie*. De Gruyter, Berlin.

Rinnooy Kan, A. H. G. (2012). Ethiek en or. *STAtOR*, 3-4:21.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15(3):351–357.

Sejnowski, T. J. (2018). *The Deep Learning Revolution*. MIT Press.

Shannon, C. E. (1950). Programming a computer for playing chess. *Philosophical Magazine*, 7(41):314.

Snelders, H. (2012). Negentiende-eeuwse theorieën over de materie. *GEWINA/TGGNWT*, 4(4):168–187.

Spencer, H. (1879). *The Principles of Ethics*. Liberty Fund, Indianapolis.

Starmans, R. J. C. M. (2011a). *Targeted Learning: Causal Inference for Observational and Experimental Data*, chapter Models, Inference and Truth: Probabilistic Reasoning in the Information Era. Springer Verlag.

Starmans, R. J. C. M. (2011b). Wikipedia, Ariane's thread or the devolution of Diderot's ideal (in Dutch). *Filosofie, Tweemaandelijks Vlaams-Nederlands Tijdschrift*, 21(3).

Starmans, R. J. C. M. (2015). Contemporary dystopias: from Icarus' fall to the wrath of the machines (in Dutch). *Filosofie, Tweemaandelijks Vlaams-Nederlands Tijdschrift*, 25(1).

Starmans, R. J. C. M. (2018a). Along the caves of morality: about ethics, statistics and data science (in Dutch). *STAtOR*, 18(1).

Starmans, R. J. C. M. (2018b). "And further away their evening-red storms and falls": on Newton, Goethe and the Light (in Dutch). *Filosofie, Tweemaandelijks Vlaams-Nederlands Tijdschrift*, 28(1).

Starmans, R. J. C. M. (2018c). A contemporary euthyphro dilemma: on deep learning and the columns of oracular language (in Dutch). *Filosofie, Tweemaandelijks Vlaams-Nederlands Tijdschrift*, 28(3).

Starmans, R. J. C. M. (2018d). Expedition robinson: from ecological correlation toward multi level analysis. *STAtOR*, 19(2).

Starmans, R. J. C. M. (2018e). Statistics and causality: progress of a laborious dialogue (in Dutch). *STAtOR*, 19(4).

Starmans, R. J. C. M. (2018f). *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*, chapter The Predicament of Truth: on Statistics, Causality, Physics and the Philosophy of Science. Springer Verlag. Springer Series in Statistics.

Starmans, R. J. C. M. (2019a). Beyond apology: theory of probability and fallible thinking (in Dutch). *STAtOR*, 20(1).

Starmans, R. J. C. M. (2019b). Cause and effect: considerations in contemporary thinking about causality (in Dutch). *Filosofie, Tweemaandelijks Vlaams-Nederlands Tijdschrift*, 29(3).

Stiegler, B. (1994). *La technique et le temps 1. La faute d'Épiméthéé*. Galilée, Paris.

Stiegler, B. (2014). *Per toeval filosoferen: een verzameling uitgeschreven radiointerwiews met Stiegler*. Klement/Pelckmans. Translated by Pieter Lemmens.

Tacq, J. (2011). Causality in qualitative and quantitative research. *Quality and Quantity*, 45(2):263–291.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59:433–460.

Tversky, A. and Kahneman, D. (1974). *Judgement under Uncertainty: Heuristics and Biases*. Cambridge University Press.

van der Laan, M. J. and Rose, S., editors (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer Series in Statistics. Springer Verlag.

van der Laan, M. J. and Rose, S., editors (2018). *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*. Springer Series in Statistics. Springer Verlag.

Virgil (2009). *Georgics*. Oxford World's Classics.

Waldmann, M. (2017). *The Oxford Handbook on Causal Reasoning*. Oxford University Press.

Wieringa, R. (2015). *Design Science*. Springer Verlag.

Williamson, J. (2009). *The Oxford Handbook of Causation*, chapter Probabilistic Theories of Causality, pages 185–212. Oxford University Press.