

VOL. 2, NO. 3, 2020, 65–83

WHAT IS DATA AND WHAT CAN IT BE USED FOR? KEY QUESTIONS IN THE AGE OF BURGEONING DATA-ESSENTIALISM

Jakob Svensson* & Oriol Poveda Guillen **

ABSTRACT

In this article we describe the rise of a data orthodoxy that we suggest to label 'data-essentialism'. We question this data-essentialism by problematizing its premises, and unveil its ideological indebtedness to deeper (previous) currents in Western thought and history. Data-essentialism is the assumption that data is the essence of basically everything, and thus provides the ideological underpinnings for the imagination of creating an Artificial Intelligence (AI) that would transform the human race and our existence. The imagination of data as an essence is in contrast to, while often conflated with, ideas of data as traces we leave behind existing in highly connected societies. This confusion over what data is, and can be used for, underlines the importance to engage in questions of the nature of data, whether everything in the universe can be described in terms of data and the implications of subscribing to such a data-essentialist worldview. We connect data-essentialism to a revival of positivism, critique a belief in the objectivity of data and that predictions based on data correlations can be fully accurate. We end the article with a discussion of how some aspects of AI rely on data-essentialist accounts and how these have a history and roots in Modernity.

Keywords: Algorithms, Artificial Intelligence, Data, Essentialism, Modernity, Positivism, Predictions

* Malmö University, Sweden.

** Freelance.

1 INTRODUCTION

Data is on the agenda today. So-called data forms the bedrock of modern policy decisions, underlies protocols of medical health, is the basis of investment strategies, informs our knowledge of the world (see Gitelman and Jackson 2013, p. 1), influences how we see ourselves and others (according to Kennedy 2016, p. 48), acts upon us (see O’Neill 2016), and thus shapes possibilities for action (according to Bowker 2013, p. 168). Hence, today there is no doubt that we are made subjects to data (see Gitelman and Jackson 2013, p. 2), determined by our *data exhausts*, an invisible ether of ones and zeroes upon which the world increasingly depends (see Jarzombek 2016, preface p.ix). With the rise of digital tech giants such as Google and Facebook, more and more aspects of our lives are mediated by their platforms as ever-increasing amounts of information are being compiled about our consumption habits, social networks and locations. According to Jarzombek (2016, preface p. x), data becomes our new *oxygen*, or should we rather say carbon *dioxide* as a growing share of our lives are dedicated to its release, capturing and processing. As hostages of these digital tech giants, we are turned into collaborators in the creation of data surpluses (see Jarzombek 2016, p. 42). But surprisingly we seem to sympathize with our captors as we participate in these practices freely (hence the title of Jarzombek’s book: *Digital Stockholm Syndrome*). Because data lays out the promise of a more convenient and efficient future in which data processing algorithms know us users (customers) better than ourselves. This is nicely illustrated in a quote by an anonymous Facebook user: ‘I am never quite sure if Facebook’s advertising algorithms know nothing about me or more that I can admit to myself’ (in Andersson Schwarz 2018, p. 68). In other words, data measuring technologies have become ingrained in the experience of the self as also the whole *Quantified Self* community is an example of.¹

It is therefore not surprising that debates over data— how it is produced, who owns it and has access to it, and to what uses it can be put — have become key political discussions in our time. The scandal with Cambridge Analytica, browsing millions of Facebook profiles and using their data traces without consent and used for political purposes in elections, is a case in point. Still, the amount of data currently harvested and its implications for our daily lives will be negligible in comparison to what the internet of things aims to deliver in terms of all-round connectivity and data-harvesting (see Bunz and Meikle 2018). Social media giants harvesting of enormous amounts of user data (as imperative for their business models)

¹ A community of experimenters in self-tracking technologies hoping that through smarter machines and their more intimate and persistent measuring, they will reach a higher degree of self-knowing.

has awakened fears of a dystopian world in which surveillance and control by a digital ‘big brother’ would offer the ultimate oppressive tool for any authoritarian regime. China is already enforcing a system of mass surveillance and control using facial recognition and big data analysis technology in their so-called *Social Credit* system which are used for among other things decisions on banking credits, insurance premiums and possibilities for travelling abroad.² Such oppressive uses of data have led researchers to address issues of data justice, relating data-driven forms of governance to broader social justice agendas (see Dencik, Hintz and Cable 2016; O’Neill 2016; Noble 2018).

In other words, we seem to be poised at the cusp of a data revolution, which makes reflections about data – what it is and what it can be used for – all the more important. However, the nature and materiality of data is seldom attended. In order to initiate a discussion about these questions we will describe what we have observed as a rise of a data orthodoxy that we suggest to label ‘data-essentialism’. This is different from the common conception of data is as *traces* we leave behind, or *exhaust* as Jarzombek (2016) would phrase it. Data-essentialism, in contrast, is based on the assumption that data is the *essence* of basically everything. An example of such data-essentialist reasoning is when acclaimed historian Harari (2015) suggests that all organisms (including humans) *consist* of data flows.³ This idea – that the building blocks of *both* computers and organisms are data – makes the merging of life sciences and data sciences possible, providing the ideological underpinnings for the belief that the human brain can be accurately modelled in a computer (see the *Human Brain Project* funded by the European Union, <https://www.humanbrainproject.eu/>)⁴ This opens a possibility for creating a form of AI that in the end would make the human race as we know it come to an end (for accounts of such scenarios, see Bostrom 2014; Tegmark 2017). Barriers between animals and machines collapse and the expectation is that electronic algorithms will decipher and eventually outperform biochemical ones (as Harari 2015, p. 428 argues). Harari even seems to claim that we already have the amount of data available, and the processing power, to upgrade our old algorithmic processor (i.e. the body).⁵ Homo Sapiens is on the brink of evolving into a

² See https://en.wikipedia.org/wiki/Social_Credit_System, accessed November 27th 2019.

³ It is unclear in his account whether the data he argues flows through human bodies are inherent to our bodies or external (or a mix of the two).

⁴ Such thinking can be found in early Cybernetics in which communication and messages are considered the backbone of *both* animals and machines (see Wiener 1948). He compared the nervous system with the computing machine of his era (see p.14).

⁵ We choose the term “seems” here as Harari (2015) in other parts of his book is ambivalent about the ability of technology to eventually making Homo Sapiens obsolete (see p. 458).

new species; *Homo Deus* (as is the title of Harari's book), or should we rather say *Homo Datus*?⁶

We believe that it is important to tease out how approaches to data evolve and differ in order to have an informed discussion about data and its power and politics in contemporary connected societies. Because these two views – data as essence and data as traces – are sometimes conflated. In this article we will attempt to distinguish the two by defining data-essentialism along three tenets (or beliefs) in which it differs from perceiving of data as traces; 1) that everything in the universe can be understood as data, 2) that data provides an objective picture of humans and hence 3) also may predict the future accurately. We will also critique and problematize these premises and link data-essentialism to a revival of positivism. We will end the article with unveiling its ideological indebtedness to deeper (previous) currents in Western thought and history. Accounts of superhuman AI (see Bostrom 2014; Tegmark 2017), rests not only on the assumption that humans can be reduced to data, but also on older assumptions inherited from Modernity that humans can be reduced to their minds.⁷

Rather than a coherent movement of people, data-essentialism is a way for us to illustrate how conceptions of data differ and sometimes are conflated. This confusion of what data is in contemporary accounts, became apparent when reading Cheney-Lippold's (2017) book with the rather misleading title *We are data*. Cheney-Lippold (2017) claims that we are 'made of data' (preface p. xiii) and that we are 'filled with data' (p. 3). However, it would be wrong to label Cheney-Lippold a data-essentialist. Reading the book to the end, the main message is actually that we are *not* made of data, but rather represented, categorized and regulated by data, and that data-mining and triangulating processes are increasingly automated without our direct participation. But to be acted upon by data, algorithms and automated systems, is not the same thing as to be made of data and this we will argue has important implications on what data can be used for.

2 DATA-ESSENTIALISM AND ITS THREE TENETS

While we rather subscribe to an approach to data as traces we leave behind living in societies permeated with digital technologies, data-essentialism assumes that we are made of data/outcomes of algorithmic calculations on data-flows. One example of such data-essentialist reasoning is in Harari's

⁶ or *Homo Sapiens Digital* as Prensky (2009) suggests.

⁷ An argument that we no longer live in Modernity but in the Global Age can be found in Poveda and Svensson 2016.

(2015) *Homo Deus*. Here he argues that human feelings are supposed to be outcomes of calculations of data in our bodies (p. 97). Free will is just biochemical processes of calculating data in order to make decisions based on probabilities (p. 328). Another example of data-essentialist thinking is Andersson's (2008) (in)famous account of the '*end of theory*'. Academics won't need theories as we have enough data and smart enough data-calculating algorithms to find patterns and hypotheses without the guidance of human thinking. Powerful computers equipped with such algorithms will be able to mine big datasets for patterns revealing effects without experimentation (as also Prensky 2009 argues), exposing patterns and relationships we didn't even know existed (see Dyche 2012), correlations that provide a full resolution of the world (see Steadman 2013), freed from human bias and framing, transcending context and thus being inherently truthful.⁸ The scientist's role shifts from being proactive (suggesting theories) to reactive with algorithms doing all of the contextual work (as Steadman 2013 forecasts). This is about collecting data first and later let the algorithms ask the questions (see Croll 2012). Such data-essentialist thinking can of course be questioned. But before this we need to better understand what data-essentialist thinking consists of.

We have identified three tenets upon which data-essentialism rests and also differs from an understanding of data as traces. The first one is the belief that everything can be accurately described in terms of data flows. One example here is Harari (2015) who argues that the wall between the organic and the inorganic has been dismantled, "turning the computer revolution from a purely mechanical affair into a biological cataclysm" (p.402). He therefore concludes that the human body is a data processing system, an algorithm (see also Wiener 1948) with everything from human imagination and feelings to free will being a product of biochemical algorithms processing data in our bodies. Neurologists have convincingly argued that the brain indeed does process information from our body which then could be behind feelings, emotions and consciousness (see for example Damasio 1999). But that such information comes in the form of data (whatever data is supposed to be in these accounts), and whether its processing is following strict steps defined in an algorithmic formula, remains questionable.

The second tenet is the imagination that it would be technically possible to extract and make calculations upon the data that our bodies is supposed to consist of. This is the belief that algorithms and automated systems may arrive at insights by correlating data being extracted from us into patterns (as Brooks 2013 seems to argue) and provide a complete and objective picture of human beings as well as a full resolution on the social

⁸ See also Kitchin's (2014, p. 132) critique of such faith in data.

worlds and cultures we humans organize ourselves in (see Steadman 2013). Another example is Harari (2015) who claims that with the rise of bio-metric devices (DNA scans et cetera), Google and its competitors will become an “all-knowing medical health service” (p. 392). In effect, this means that a human being can be reduced (from her bio-chemical processes to her social behaviour) to the data extracted from her in what is supposed to be a scientific and bias-free way. Others have described this as *dataism*; a “widespread belief in the objective quantification and potential tracking of all kinds of human behaviour and sociality through online media technologies” (van Dijck 2014, p. 198). This in turn is linked to *datafication*, the paradigm for understanding sociality and social behaviour by transforming social action into online quantified data (see Cukier and Mayer-Schoenberger 2013).

This leads us to the third and final tenet of data-essentialism: By compiling and analysing increasing amounts of data harvested from human beings, it is believed to be possible to make *fully accurate predictions* about our behaviour. That data traces we leave behind can tell us great deal, seems like an uncontroversial claim. But only if we imagine data as neutral and objectively true may it allow for fully accurate predictions. Indeed, objective quantification and tracking is only possible if data is conceived of as a *neutral essence* rather than as *contextual* and *situated traces*. Taken to its logical consequence, what this third tenet postulates, is that with enough data, predictions would no longer be a matter of probabilities but would rather evolve into error-free forecasts. One example of how such thinking can have potentially harmful consequences is so-called predictive policing. While being presented as objective and bias-free, O’Neill (2016) shows how predictive policing systems send cops back to the same poor neighbourhoods, creating a toxic feedback loop since policing one street creates new data that justifies more policing in that exactly that same street. As Siegel (2013, p. 90) claims, we do not need to care about causation, explaining *the why*, when the objective is to predict the world rather than to understand it (an argument Pearl and Mackenzie 2018 refutes in their *Book of Why*).

There is no doubt that data and its processing by algorithms have wide ranging implications in terms of how we are represented, controlled and disciplined today (as O’Neill 2016, Cheney-Lippold 2017 and Noble 2018 have shown). Life in connected societies indeed increasingly takes place in and through an algorithmic media landscape processing data (as Bucher 2018 argues). We are datafied, including our friendships and relationships (see Kennedy 2016, p.10). But this is not the same as consisting of data, or that data is a neutral essence. But this begs the question of what data really is, which leads us to the next section.

3 WHAT IS DATA?

Kitchin (2014, p.18) claims that data is getting an ontological status in technology, sociology as well as biology. At the same time, he complains that little attention has been paid to data's ontological framing and the meaning of data itself (p. 25). We agree that data is often treated as ontological, but that questions about its nature and materiality remain unanswered. Cheney-Lippold (2017) does not define data despite the title of his book. Harari (2015) uses the terms *data* and *information* interchangeably without defining neither of them. This lack of definition is arguably behind confusions of what data really is, and for some researcher to treat data as a neutral essence, devoid of cultural bias.

The treatment of data as an objective and neutral reflection of reality resonates in the etymological meaning of the word as something that is given (from *datum* and the Latin verb *dare* i.e. to give, see Rosenberg 2013, p. 18). In 17th century philosophy, data equalled facts and principles that were “by agreement beyond argument” (Rosenberg 2013, p. 20). Here data is supposed to be the starting point of what we know and cannot be deconstructed. This etymological meaning is probably behind conceptions of data as an essence, “transparent, autonomous, objective and neutral” (Gitelman and Jackson 2013, pp. 2-3). However, data is not given, most often it is *captured*, extracted through observations and computation.⁹ But even though the meaning of data has shifted from the rhetorical (what is beyond argument), to the observable (what can be extracted, see Rosenberg 2013, p. 36), its connotation to the objective and factual seems to have persevered. Data is supposed to have no inherent meaning (as Kitchin 2014, p. 17 argues), and therefore it has been very useful as a concept (according to Rosenberg 2013, p. 37).

There are critics of perceiving of data as objective. Data do not just exist; it has to be generated. This is nicely illustrated in Ribes and Jackson's (2013) study of the largely invisible infrastructures of data, how scientists and technicians worked hard to make data the same, comparable over a long period of time in a setting in which context and conditions were constantly changing. Etymologically it would make more sense to talk about data as *information*. Galloway (2011, p. 87) distinguishes *what is given* from *information*, meaning the act of *being formed* or *put into a form*. Hence, the data that is most often referred to today, is *computable* data, data that is made ‘algorithm-ready’ (Bucher 2018, p. 5), ‘scrubbed’ (Gitelman and Jackson 2013, p. 7) and ‘cleaned’ (Kennedy 2016, p.108) for computer algorithms to use in their calculations. And computer-readied data is not

⁹ Which leads Kitchin 2014, page 2, to suggest we should rather talk about *capta* rather than *data* (from Latin *capere* i.e. to capture).

formless. It is captured when being measured/collected, a capture which shapes the data (see Ribes and Jackson 2013) and put into a quantified form of ones and zeroes in order for computers to process it (see Kennedy 2016, p. 10). Data is thus always dependent on developments around its capturing and scrubbing (Pink et al. 2018).

This suggests that data is deeply cultural and infused with societal norms and values. Data does not naturally appear as it is collected and manipulated by people, shaped by human decisions, interpretations and filters (see Kennedy 2016, p. 110; Cheney-Lippold 2017, p. preface xiii). Behind data production there are assemblages of people, places, documents, practices and technologies, making data a product of complex processes in order to be useful for the contexts in which it appears (as Ribes and Jackson 2013 show). Krippendorf (2016) therefore defines data as a *human artifact*. Indeed, data is both social (situated in a context), material in that it has a form. In terms of computer data this would be in the form of bits stored on a hard drive, and depending on infrastructures (such as data centres and cables, see Holt and Vonderau 2015). Raw data is thus ‘an oxymoron’ (as Gitelman and Jackson 2013 argue) and should be ‘cooked with care’ (Bowker 2005, p. 184), otherwise it might ‘rot’ (Boellstorff 2013) and thus be in need of ‘repair’ (Pink et al. 2018).

This reasoning above is surprisingly uncontroversial. There is even a field called *critical data studies* (see Illiadis and Russo 2016). In tech literature such as *Algorithms for Dummies* (Mueller and Massaron 2017) it is clearly stated that data is not raw, it is managed and that programmers and algorithms are so-called ‘data managers’ (see p. 68). Once we take away the neutrality and objectivity of data, admit that it is socio-cultural, the question is if the premises upon which ideas of superhumans and AI rest also would start to unravel? Because, if we agree that data is a human construct, how could everything in the universe be described in terms of data? Or is data-essentialism an extreme form of social constructionism?

Here it seems that data-essentialism is connected to the hype around so-called *big data* (see Cukier and Mayer-Schoenberger 2013). This is the imagination that large datasets open the possibility for a higher form of intelligence and knowledge and thus may generate insights that were previously unavailable through hidden patterns and correlations in data points (i.e. data-essentialism’s 2nd tenet). Andrejevic (2020, p. 35) talks about this as a *fantasy of framelessness*. Automated collection and processing of data is thought of as final and ultimate, as it nurtures a *fantasy of total information collection* (Andrejevic 2020, p. 35) out of which decision untouched by human prejudices can be made.

Big data is the outcome of an increasing ease and thus intensification of data collection and storage coupled with computers with increased processing power. Digital storage solutions have reduced the cost and space

of retaining data, and the networking of computers has facilitated the transfer and sharing of data (see Kitchin 2014, pp. 31, 82). However, big data is a relative term. Big data is only big in relation to previous amount of data collection and processing. It is indeed big compared to what human beings alone can process, but it is small compared to the amount of data potentially available (see Poveda and Svensson 2016). It is therefore important not to confuse *big* data with *all* data (as also Andrejevic 2020 argues). Data harvested through measurement is always a selection from the total sum of all possible data (see also Kitchin 2014, p. 3). And since so-called big data cannot capture the whole picture (it is always framed), calculations on big data sets are biased from the beginning as they constitute partial orders, localized totalities and with an ability to only gaze in some directions but not others (see Kitchin 2014, p. 133). As Cukier and Mayer-Schoenberger (2013) reminds us of, “however dazzling the power of big data appears, its seductive glimmer must never blind us to its inherent imperfections” (p. 28).

4 WHAT CAN DATA BE USED FOR?

If we agree that data is (inter)subjective, infused by socio-cultural norms and values (at least in part), we should also start to ask what it can be used for. In an interview with a software engineer he stated that “data you do not do anything with, is uninteresting”, that “data can be bad and not useful” and that “data only treats one part of reality” (in Svensson 2020). Hence, if we know that data from the beginning is biased, that big data is far from all data, how can the predictions it makes be fully accurate and applicable? Furthermore, algorithms are trained to find correlations in data, make associations and construct patterns and out of these make predictions out of probabilities. Patterns need big numbers and thus mostly work on big data sets. It is by collecting *enough* data that not only the past and present are mapped, but also the future. And the more the coin is flipped, the more the result will converge upon the precalculated probability (see Steiner 2012). Indeed, patterns are all about prediction which is all about probabilities. Already Wiener (1947, p. 34) was occupied with the ability to predict out of information. This fascination with prediction goes all the way back to Leibniz who thought humans were programmed to behave in certain manners (according to Steiner 2013, p. 61). But correlation does not supersede causation and data does not understand causes and effects (as Pearl and Mackenzie 2018, p. 21 argue). As Cukier and Mayer-Schoenberger (2013) states, the use of big data might imply we will need to give up our quest to discover the cause of things. Looking for patterns might help predict the future, answer to what probably (but not certainly) will happen, but not why this will happen. Hence, predictions are only probabilities and

are not always correct. And as Pearl and Mackenzie (2018, p.47) argue, causation is not reducible to probabilities. Even if predictions would be based on completely neutral and correct data, the people using these systems might not be, as the case of predictive policing has shown. It becomes dangerous if we treat predictions of probabilities as undeniable truths.

Since algorithms will not ask why they get the results they get or what the consequences of their results might be, it makes them blind to ethical issues (see Diakopoulos 2016). This is about outsourcing the ordering of the world we inhabit to algorithms lacking reflexive capabilities and lacking agency to handle the messiness of the present (see Klinger and Svensson 2018). Hence, there are numerous examples of when algorithms fail, such as Amazon being accused of homophobia (see Striphas 2015), Google of racism (see Noble 2018), gender biases of image-search algorithms (see Kay, Matuszek and Munson 2015) and cases where black people are not recognized as humans in face-recognition algorithms (see Sandvig et al. 2016).

It is only if we believe in the objectivity of data, imagine that it would be technically possible to extract and make calculations upon the data in our bodies, that patterns found in big datasets could be used for fully accurate predictions. But if we believe that data are traces that we leave behind in a digital existence, such predictions would always be based in the past. This contemporary craving for patterns may have dire consequences when making judgements about people's ability to change destructive patterns of the past (see O'Neill 2016). The past is not necessarily determinative of the future, people can change. If we instead approach data as traces from past behaviour online, algorithmically calculated patterns, these cannot be believed to predict the future with 100 percent of accuracy. If there is something we have learned in the history of humankind, is that it has taken many unexpected turns.

To be human is to be random, unfinished, imperfect and disorderly, to be a constant "beta version" (as Cheney-Lippold 2017, p. 90, eloquently puts it). At the same time, most of data analytics and processing are about orderliness, calculations and finding patterns which are supposed to predict future behaviour. A software engineer interviewed actually described code as a grammar with no exceptions (see Svensson 2020). This was the reason why he loved coding, comparing this to struggles with German grammar at high school. But as humans we have plenty of exceptions and at times we act randomly and in a surprising manner. As Morozov (2013, introduction p. xiii) argues, sometimes imperfect is good enough and even much better than perfect. It seems that the orderliness of programming and code languages, are at odds with human imperfectness and randomness. Maybe some things are just un-representable by

computer-readied data in the form of ones and zeros (as also Galloway 2011 argues). Maybe this is why we sometimes feel creeped out by our datafied selves (see Cheney-Lippold 2017, p. 193). We are recognizable but in an odd way. It becomes uncanny in the same way that robots can be creepily similar, but not quite like the real thing (the so-called uncanny valley).¹⁰ Behind the perfect surface, there is just mechanical impulses. Digital computers can mimic the actions of human behaviour as already Turing (1950, p. 437) forecasted. But is an imitation the real thing? Arguably what is missing in our datafied replicas/upgrades is irrationality and randomness, patterns and also correlations, but with plenty of exceptions.

5 A REVIVAL OF POSITIVISM?

The belief that data can capture everything with full resolution, freed from human bias, framing and context does ring a bell. The bringing of the unruly social world into the formal study of the natural sciences, rendering culture and society computable is surrounded by a discourse of positivistic measurement. It thus seems data-essentialism is accompanied with a revival of positivism within the Social Sciences. Indeed, as Kitchin (2014, pp. 139-140) argues, data-driven sciences favour transforming research about humans and their societies to something resembling natural and engineering sciences, offering opportunities for a ‘truthful’ study of human life. Törnberg (2019) labels the use of API-based technologies to inductively seek patterns as *predicative positivism*. Indeed, datafication implies transforming sociality, behaviour and culture into *quantified data* to be used for real-time tracking and predictive analysis.

Following this discourse of positivistic measurement, Anderson (2008) has (in)famously argued that theory has come to an end, and that we now have enough data and fast enough computers to actually study *the physics of culture*. He thus seems to suggest that data will be able to speak for itself. This can of course be questioned (see also Törnberg 2019). Bucher (2018, p. 24) for example claims that without algorithms, data would just flow without any particular direction. Algorithms are actually an outcome of media logics rather than a replacement of them (see Klinger and Svensson 2018 for an outline of this argument). Algorithms are based on hypotheses from the beginning (see Bucher 2018, p. 25). And even if data harvested from social media platforms is supposed to reflect human behaviour, the algorithms employed (by Google, Facebook and others) are intrinsically selective and manipulative to suit the interests of these companies (see van Dijck 2014, p. 200). Hence, it is easy to dismiss statements such as those claiming that data speaking for itself. But it is

¹⁰ See https://en.wikipedia.org/wiki/Uncanny_valley, accessed August 21th 2020.

important to understand that one of the strongest epistemic conditions shaping data imaginaries today, is the self-evidence of numbers. Data's connection to numbers and mathematical functions, gives it an allure of neutrality and objectivity, which in turn makes humans look particularly subjective and biased in comparison (see Bucher 2018, p. 56).

Kennedy (2016, p. 150) talks about a 'pervasive desire for numbers' as an emerging rationality today. She shows in her studies of public sector organizations that mere numbers are met with enthusiasm (even though it was not always clear what they stood for). Kennedy connects this desire to earlier studies about trust in numbers that seem to support the prestige and power of quantitative methods. Numbers can be understood from far away and are universal as they can be shared across cultures (see Kennedy 2016, p. 81). They are impersonal, therefore also appear to be objective and thus credible. This to the point that even friendships and sociality are quantified in a positivistic manner of objective measurement (see Bucher 2018, p. 9). What was once qualitative has been turned into numbers. According to Kennedy (2016, pp. 100-101), this limits the possibility to discuss the ways in which data is made and shaped.

This desire for numbers, with its allure of objectivity and neutrality, is accompanied with a belief of unbiased calculation, the translation of everything into mathematical symbolic language following mathematical laws. Algorithms introduce and privilege quantification and automation, the ordering of various types, statistical reasoning and large numbers (see Bucher 2018, pp. 31-32). And if we are made up of data and our bodies are just bio-chemical algorithms processing this data, this also means that we humans could be fully predicted in mathematical formulas, that the entirety of our everyday life practices and ourselves are subject to – and constituted by – perpetual calculation (as Raley 2013, p. 126 argues). Harari (2015, p. 99) gives the example of a baboon spotting some ripe bananas in-between him and a lion. His body will calculate how hungry he is together with probability of success, which will then result in a feeling of bravery or caution. In other words, sensations, emotions and actions are a result of mathematical calculations on the data inside of us according to Harari (2015, p. 124). Harari (2015, p. 101) even argues that attraction and beauty are results of years of calculating data about reproduction with successful offspring. But is it really possible to reduce subjective and intersubjective experiences such as beauty to mathematical calculations on data? If it is one thing we know about beauty, it is that it is culture specific, whereas today's Western beauty ideal of female skinniness is not related to being successful at birth-giving (arguably it is the other way around). Indeed, as Bucher (2018, p. 11) puts it, by reducing human connections to algorithmic calculations, we risk dehumanizing sociality. People are not a math problem, and people are more complicated than an equation, more complex

and unpredictable than what can be broken down into a few steps of instructions in a computer algorithm (as Bucher 2018, pp. 104-105, argues).

6 AI AND ITS ROOTS IN MODERNITY

By reducing us humans, our connections and our behaviour to data being algorithmically processed, calculated by our bodies or/and computers, data-essentialism provides the ideological underpinnings for the belief that humans can be replaced by AI with far greater capabilities (see for example Bostrom 2014; Harari 2015; Tegmark 2017). According to this line of reasoning, it would be technically possible to create machines that are better and more efficient at processing our data. This claim is currently challenged by science's poor understanding of how human consciousness works (see Damasio 1999). But this might be a temporary obstacle that new research perhaps could contribute to overcome.

A more problematic objection can be found in AI's understanding of the human. It is worthwhile to interrogate in which ways the reduction of being human to data is indebted to older forms of reductionism. In religious thought, the search for a human essence detached from the physical body led to the notion of an immortal soul. In modern times, Descartes (2017) gave scientific sanction to the body/soul dualism previously upheld by theologians by reframing it as the body/mind split. Descartes, too, conceived of bodies as machines. Data-essentialism reproduces in a magnified fashion the soul/body controversy in Christianity. The project to de-incarnate the human and retrieve her essence, has ancient roots but current discussions around AI seem not to account for this ideological lineage and presents it as novel, what is in fact a cultural bias with a long history in Western thinking.

It is worthwhile to look past the hype that surrounds AI and to question its claim for novelty. As a matter of fact, data-essentialism's first and second tenets were already expressed by Weber (2008) in his famous lecture series when he described *disenchantment* as "the knowledge or belief that if we only wanted to, we could learn at any time that there are, in principle, no mysterious unpredictable forces in play, but that *all things — in principle — can be controlled through calculation*" (p. 35, emphasis in the original). The third tenet of data-essentialism, the belief that with enough data it would be possible to make fully accurate predictions, seems also to be behind Weber's reference to absolute control. As much as data-essentialism toys with the idea of rendering humans obsolete, it is important to underline that, historically speaking, the modern project of human mastery lies at its core. The modern belief in endless progress lurks behind ideas upgrading humans with computer technology. Indeed, As Morozov (2013, introduction p. ix) argues, to question Silicon Valley's quest

to solve any kind of problems with tech, has become equivalent of questioning Enlightenment itself. Also, Rosenberg (2013, p. 15) associates the rise of the concept of data to Modernity and Jarzombek (2016, p. 39) argues that data processing is about making the Self and Others predictable, identifiable and exploitable. To participate in the project of Modernity has always meant that one becomes “a calculable subject” (according to Raley 2013, p.126). And what is the meaning of AI apart from *progress* and a trust in an upgraded future? However, even if we would consist of data flows, it would still be uncertain that we would process this data in a *rational* manner. The modern belief in rationality, that human beings act (at least in the aggregate) as rational beings and in their self-interest is part of AI. However, global warming clearly shows otherwise. For the sake of ourselves and our survival, the most rational thing to do would be to reduce our carbon footprints (while on the contrary, it seems to be increasing). Indeed, the façade that attempts to present AI as a dispassionate reckoning with the objective realities of today, our data and algorithm saturated world belies a much more complicated and problematic genealogy of its foundational principles and ideas.

Finally, it is relevant to point out that our critique of data-essentialism is not predicated upon any form of human exceptionalism. Intelligence and a rich emotional life are not an exclusive prerogative of human animals. Our critique aims rather at problematizing the premises of data-essentialism and to unveil its ideological indebtedness to deeper currents in Western thought and history that have little to do with claims of objectivity and neutrality.

7 CONCLUDING REMARKS

Many people today believe in data as we ask Google and Facebook for advices on a range of different matters. Contemporary life is indeed characterized by data collection and processing. As we are *thrown* into a digital existence (see Lagerkvist 2017), digital tech giants and data scientists are increasingly powerful centres around which our existence gravitate. But acknowledging the importance of data, conceiving of data as contextual and situated traces we leave behind in an increasingly computer saturated world is substantially different from reducing our existence and bodies to data. As we have discussed in this article, such data-essentialism is indebted to modernist thinking about progress, calculation and rationality.

Harari (2015, p. 207) does emphasize the role of fiction for societies to function. The importance of SciFi (Science Fiction) in tech in general and AI in particular cannot be understated. SciFi aesthetics, with its connection to futurism, are all over tech culture (see Svensson 2020). The modern imagination of a disembodied future also resonates in SciFi classics such as

Gibson's (1984) *Neuromancer*. At the 2019 South by Southwest festival Cassie Kozyrkov, chief data scientist at Google, argued that the only reason AI got funding in its early days was because of its appeal to SciFi. Similarly, data-essentialism seems to be based on a powerful modern fiction of humans as rational, predictable and therefore also controllable.

Bowker (2013 p. 171) writes that computers may have data, but that not everything in the world is given. Indeed, it makes more sense to understand data as partial translations (as every translation is partial, imperfect) of perceived reality in mathematical language. As such, data-essentialism seem to suffer from a poor understanding of semiotics as they mix up the sign with the thing itself. In this sense, data-essentialism is social constructionism trapped inside a cage of mathematical language, which, by virtue of being more abstract than regular human language, appears to be purer or even divinely inspired (as in Harari' 2015 account of *Homo Deus*). Data is not only a representation; it is also always a sample. Even big data is only a representation, not a totality, stand-ins for phenomena of theoretical or practical importance (see Krippendorff 2016). And to base our whole being, existence and future on partial data-traces we leave behind in mathematical language, on the "residues of human existence in a digital world" (Cheney-Lippold 2017, p. 89), would be akin to a *synecdoche*, to take a small piece and make it a representative sign of a totality.

As AI is developing now, there is no reason to believe it can fully replicate humans any time soon. Today AI is only executive while humans also think creatively and have a reflective character (see Hindi 2017). Data processing machines can show emotions but not feel them and this is different. Even a data-enthusiasts such as Domingos (2015) state that only because computers can learn "they will not magically acquire a will of their own" (p. 45). Case (2018) therefore argues that humans *together with* AI (something Case labels as *centaurs*) seem to be a winning team (even against teams of computers only). So, it seems that intelligence is not a single dimension, and that human intelligence includes random, creative, unruly and scattered elements that are hard to capture in algorithms processing readied/cleaned data.

Turing (1950, p. 440) with his focus on imitation and mimic, suggested a clear hierarchy from the human to the machine. As a gay man in the UK during the World War 2, he knew what it was like having to pass as a straight man. Today transgender activist Vanessa López raises questions about what it takes to pass as a woman in a Western society (see her book from 2014 about her regretting her gender reassignment surgery). In a similar manner we could ask whether we cannot let artificial intelligence be artificial intelligence? Does it have to *pass* as human? Why this pre-occupation with *passing* within AI? We should instead focus on what machines and AI are good at and what humans are good at, and how we

together can be at the service in relation to the big problems we as humans and our planet are facing, such as xenophobia, polarization, intolerance and climate change.

FUNDING STATEMENT (AND ACKNOWLEDGMENTS)

We want to recognize the generous funding from the Swedish Research Council as well as the fellowship program at Weizenbaum Institute for the Networked Society, making this research possible. Furthermore, we would like to acknowledge colleagues at Malmö University providing important feedback to our ideas at an early stage, as well as this Journal's excellent reviewers.

REFERENCES

- Anderson, C. (2008). End of Theory. The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine*, 23 June 23.
<https://www.wired.com/2008/06/pb-theory/>, accessed 16 July 2016
- Andersson Schwarz, J. (2018). Umwelt and Individuation: Digital Signals and Technical Being. In Lagerkvist, A. (ed), *Digital Existence. Ontology, Ethics and Transcendence in Digital Culture*. New York: Routledge, pp. 61-80
- Andrejevic, M. (2020). *Automated Media*. London Routledge
- Bowker, G. C. (2005). *Memory Practices in the Sciences*. Cambridge: MIT Press
- Bowker, G. C. (2013). Data Flakes: An afterword to "Raw Data" Is and Oxymoron. In Gitelman L. (ed.), *Raw Data is an Oxymoron*. Cambridge: MIT Press, pp. 167-171
- Boellstorff, T. (2013). Making big data, in theory, *First Monday*, 18(10).
<https://firstmonday.org/article/view/4869/3750>, accessed 8 December 2017
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press
- Brooks, D. (2013). The Philosophy of Data, *New York Times*. 8 February 18.
<https://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html>, accessed 20 August 2017.
- Bucher, T. (2018). *If ... Then. Algorithmic Power and Politics*. Oxford: Oxford University Press.
- Bunz, M., and Meikle, G. (2018). *The Internet of Things*. Cambridge: Polity Press.
- Case, N. (2018). How to become a centaur. *Journal of Design & Science*,
<https://jods.mitpress.mit.edu/pub/issue3-case>, accessed 20 August 2017

- Cheney-Lippold, J. (2017). *We are Data. Algorithms and the Making of our Digital Selves*. New York: New York University Press
- Croll, A. (2012). Big data is our generation's civil rights issue and we don't know it, *O'Reilly Radar*, 2 August 2012.
<https://www.cc.gatech.edu/~beki/cs4001/big-data.pdf>, accessed 20 August 2017
- Cukier, K., and Mayer-Schoenberger, V. (2013). The Rise of Big Data. How It's Changing the Way We Think about the world, *Foreign Affairs*, 92 (3), pp. 28-40
- Damasio, A. (1999). *The Feeling of What Happens. Body and Emotion in the making of consciousness*. New York: Houghton Mifflin Harcourt Publishing Company
- Dencik, L., Hintz, A., and Cable, J. (2016). Towards data justice? The ambiguity of anti-surveillance resistance in political activism, *Big Data & Society*, July -December, pp. 1-12.
<https://journals.sagepub.com/doi/10.1177/2053951716679678>, accessed 20 August 2017
- Diakopoulos, N. (2016). Accountability in Algorithmic Decision Making. *Communications of the ACM*, 59(2), pp. 56-62
- Descartes, R. (2017). *Meditations on First Philosophy, with Selections from the Objections and Replies*. Translated and edited by John Cottingham. 2nd edition. Cambridge: Cambridge University Press. Original work published 1647
- Domingos, P. (2015). *The Master Algorithm. How the Quest for the Ultimate Learning Machine will Remake our World*. London: Penguin
- Dyche, J. (2012). Big Data Eureka's do not just happen. *Harvard Business review blog*, 20 November 2012. <https://hbr.org/2012/11/eureka-doesnt-just-happen>, accessed 20 August 2017
- Galloway, A. (2011). Are Some Things Unrepresentable? *Theory, Culture & Society*, 28(7-8), pp. 85-102
- Gibson, W. (1984). *Neuromancer*. London: Gollancz
- Gitelman, L., and Jackson, V. (2013). Introduction, In Gitelman L. (ed.), *Raw Data is an Oxymoron*. Cambridge: MIT Press, pp. 1-14
- Harari, Y.N. (2015). *Homo Deus. A Brief History of Tomorrow*. London: Vintage Books
- Hindi, R. (2017). Will Robots take over? *the Conference*, Malmö august 2017. <https://urplay.se/program/202921-ur-samtiden-the-conference-2017-kommer-robotarna-ta-over>, accessed 15 December 2017
- Holt, J., and Vonderau, P. (2015). "Where the Internet Lives": Data Centers as Cloud Infrastructures. In Parks, L., and Starosielski, N. (eds), *Signal Traffic. Critical Studies of Media Infrastructures*. Chicago: University of Illinois Press, pp. 71-93

- Illiadis, A., and Russo, F. (2016). Critical Data Studies. An Introduction, *Big Data & Society*, July-December, pp. 1-7.
<https://journals.sagepub.com/doi/abs/10.1177/2053951716674238>,
 accessed 20 August 2017
- Jarzombek, M. (2016). *Digital Stockholm Syndrome in the Post-Ontological Age*. Minneapolis: University of Minnesota Press
- Kay, M., Matuszek, C., and Munson, S.A. (2015). Unequal representation and gender stereotypes in image search results for occupations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI 15)*. New York: ACM, pp. 3819-3828
- Kennedy, H. (2016). *Post, Mine, Repeat. Social media data mining becomes ordinary*. London: Palgrave Macmillan
- Kitchin, R. (2014). *The Data Revolution. Big Data, Open Data, Data Infrastructures & Their Consequences*. London: Sage
- Klinger, U., and Svensson, J. (2018). The End of Media Logics? On Algorithms and Agency. *New Media & Society*, 20(12), pp. 4653–4670
- Krippendorf, K. (2016). Data. In *International Encyclopedia of Communication Theory and Philosophy*. New York: Wiley.
<https://onlinelibrary.wiley.com/doi/10.1002/9781118766804.wbiect104>,
 accessed 15 December 2018
- Lagerkvist, A. (2017). Existential media: Toward a theorization of digital thrownness. *New Media & Society*, 19(1), pp. 96–110
- López, V. (2014). *Jag har ångrat mig (I have changed my mind)* Stockholm: Two-Spirit Publishers
- Morozov, E. (2013). To Save everything, Click here. The folly of Technological Solutionism. New York: Public affair
- Mueller, J.P., and Massaron L. (2017). *Algorithms for Dummies*. Hoboken: John Wiley and Sons
- Noble, S.U. (2018). *Algorithms of Oppression. How Search Engines Reinforce Racism*. New York: New York University Press
- O’Neill, C. (2016). *Weapons of Math Destruction*. New York: Crown Publishing
- Pearl, J., and Mackenzie D. (2018). *The Book of Why. The New Science of Cause and Effect*. London: Penguin Books
- Pink, S., Ruckenstein, M., William R., and Duque, M. (2018). Broken data. Conceptualising data in an emerging world, *Big Data & Society*, January-June, pp. 1-13.
<https://journals.sagepub.com/doi/full/10.1177/2053951717753228>,
 accessed 8 December 2018
- Poveda, O., and Svensson, J. (2016). Re-thinking the Global Age as Interdependence, Opacity and Inertia. *Triple C*, 14(2), pp. 475-495.
- Prensky, M. (2009). H. Sapiens Digital: From Digital Immigrants and Digital Natives to Digital Wisdom. *Innovate: Journal of Online*

- Education*, 5(3).
<https://nsuworks.nova.edu/cgi/viewcontent.cgi?article=1020&context=innovate>, accessed 8 December 2018
- Raley, R. (2013). Dataveillance and Countervailance. In Gitelman L. (ed.), *Raw Data is an Oxymoron*. Cambridge: MIT Press, pp. 121-145.
- Ribes, D., and Jackson, S.J. (2013). Data Bite Mam: The Work of Sustaining a Long-Term Study. In Gitelman L. (ed.), *Raw Data is an Oxymoron*. Cambridge: MIT Press, pp. 147-166.
- Rosenberg, D. (2013). Data before the Fact. In Gitelman L. (ed.), *Raw Data is an Oxymoron*. Cambridge: MIT Press, pp. 15-40
- Sandvig, C., Hamilton, K., Karahalios, K., and Langbort, C. (2016). When the Algorithm itself Is a Racist. Diagnosing Ethical Harm in the Basic Components of Software, *International Journal of Communication*, 10, pp. 4972-4990
- Siegel, E. (2013). *Predictive Analytics*. Hoboken: Wiley
- Steadman, I. (2013). Big data and the death of the theorist. *Wired Magazine*, 25 January, <https://www.wired.co.uk/article/big-data-end-of-theory>, accessed 8 December 2018.
- Steiner, C. (2012). *Automate this. How algorithms came to rule our world*. New York: Penguin Books
- Striphas, T. (2015). Algorithmic Culture. *European Journal of Cultural Studies*, 18(4-5), pp. 395-412
- Svensson, J (2020). *Wizards of the Web. A journey into tech culture, mathemagics and the logics of programming*. Göteborg: Nordicom (forthcoming)
- Tegmark, M. (2017). *Life 3.0. Being Human in the Age of Artificial Intelligence*. New York: Vintage Books.
- Turing, A.M (1950). Computing Machinery and Intelligence. *Mind*, 49, pp. 433-460
- Törnberg, A. (2019). Teorins död? Om framväxten av en digital empirism. *Fronesis*, 64-65, pp. 132-146.
- Van Dijck, J. (2014). Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology. *Surveillance and Society*, 12(2), pp. 197-208
- Weber, M. (2008). Science and Vocation. In Dreijmanis, J. (ed), *Weber's complete writings on academic and political vocations*. New York, NY: Algora Publishing, pp. 1917-1919
- Wiener, N. (1948). *Cybernetics, or Control and Communication in the Animal and the Machine*. Cambridge: MIT Press