ARTICLE

Check for updates

Moving back to the future of big data-driven research: reflecting on the social in genomics

Melanie Goisauf [1,2,4✉], Kaya Akyüz [1,2,4✉] & Gillian M. Martin [3,4]

With the advance of genomics, specific individual conditions have received increased attention in the generation of scientific knowledge. This spans the extremes of the aim of curing genetic diseases and identifying the biological basis of social behaviour. In this development, the ways knowledge is produced have gained significant relevance, as the data-intensive search for biology/sociality associations has repercussions on doing social research and on theory. This article argues that an in-depth discussion and critical reflection on the social configurations that are inscribed in, and reproduced by genomic data-intensive research is urgently needed. This is illustrated by debating a recent case: a large-scale genome-wide association study (GWAS) on sexual orientation that suggested partial genetic basis for same-sex sexual behaviour (Ganna et al. 2019b). This case is analysed from three angles: (1) the demonstration of how, in the process of genomics research, societal relations, understandings and categorizations are used and inscribed into social phenomena and outcomes; (2) the exploration of the ways that the (big) data-driven research is constituted by increasingly moving away from theory and methodological generation of theoretical concepts that foster the understanding of societal contexts and relations (Kitchin 2014a). Big Data Soc and (3) the demonstration of how the assumption of 'free from theory' in this case does not mean free of choices made, which are themselves restricted by data that are available. In questioning how key sociological categories are incorporated in a wider scientific debate on genetic conditions and knowledge production, the article shows how underlying classification and categorizations, which are inherently social in their production, can have wide ranging implications. The conclusion cautions against the marginalization of social science in the wake of developments in data-driven research that neglect social theory, established methodology and the contextual relevance of the social environment.

[1] Department of Science and Technology Studies, University of Vienna, Vienna, Austria. [2] BBMRI-ERIC, Graz, Austria. [3] Department of Sociology, University of Malta, Msida, Malta. [4] These authors contributed equally: Melanie Goisauf, Kaya Akyüz, Gillian M. Martin. ✉email: melanie.goisauf@univie.ac.at; kaya.akyuez@univie.ac.at

Introduction

With the advance of genomic research, specific individual conditions received increased attention in scientific knowledge generation. While understanding the genetic foundations of diseases has become an important driver for the advancement of personalized medicine, the focus of interest has also expanded from disease to social behaviour. These developments are embedded in a wider discourse in science and society about the opportunities and limits of genomic research and intervention. With the emergence of the genome as a key concept for 'life itself', understandings of health and disease, responsibility and risk, and the relation between present conditions and future health outcomes have shifted, impacting also the ways in which identities are conceptualized under new genetic conditions (Novas and Rose 2000). At the same time, the growing literature of postgenomics points to evolving understandings of what 'gene' and 'environment' are (Landecker and Panofsky 2013; Fox Keller 2014; Meloni 2016). The postgenomic genome is no longer understood as merely directional and static, but rather as a complex and dynamic system that responds to its environment (Fox Keller 2015), where *the social* as part of *the environment* becomes a signal for activation or silencing of genes (Landecker 2016). At the same time, genetic engineering, prominently known as the gene-editing technology CRISPR/Cas9, has received considerable attention, but also caused concerns regarding its ethical, legal and societal implications (ELSI) and governance (Howard et al. 2018; Jasanoff and Hurlbut 2018). Taking these developments together, the big question of nature vs. nurture has taken on a new significance.

Studies which aim to reveal how biology and culture are being put in relation to each other appear frequently and pursue a genomic re-thinking of social outcomes and phenomena, such as educational attainment (Lee et al. 2018) or social stratification (Abdellaoui et al. 2019). Yet, we also witness very controversial applications of biotechnology, such as the first known case of human germline editing by He Jiankui in China, which has impacted the scientific community both as an impetus of wide protests and insecurity about the future of gene-editing and its use, but also instigated calls towards public consensus to (re-)set boundaries to what is editable (Morrison and de Saille 2019).

Against this background, we are going to debate in this article a particular case that appeared within the same timeframe as these developments: a large-scale genome-wide association study (GWAS) on sexual orientation[1], which suggested partial genetic basis for same-sex sexual behaviour (Ganna et al. 2019b). Some scientists have been claiming sexual orientation to be partly heritable and trying to identify genetic basis for sexual orientation for years (Hamer et al. 1993); however, this was the first time that genetic variants were identified as statistically significant and replicated in an independent sample. We consider this GWAS not only by questioning the ways genes are associated with "the social" within this research, but also by exploring how the complexity of the social is reduced through specific data practices in research.

The sexual orientation study also constitutes an interesting case to reflect on how knowledge is produced at a time the data-intensive search for biology/sociality associations has repercussions on doing social research and on theory (Meloni 2014). Large amounts of genomic data are needed to identify genetic variations and for finding correlations with different biological and social factors. The rise of the genome corresponds to the rise of big data as the collection and sharing of genomic data gains power with the development of big data analytics (Parry and Greenhough 2017). Growing number of correlations, e.g. in genomics of educational attainment (Lee et al. 2018; Okbay et al. 2016), are being found that are linking the genome to the social,

increasingly blurring the established biological/social divide. These could open up new ways of understanding life, and underpin the importance of culture, while, paradoxically, may also carry the risk of new genetic determinism and essentialism. The changing understanding of the now molecularised and datafied body also illustrates the changing significance of empirical research and sociology (Savage and Burrows 2007) in the era of postgenomics and 'datafication' (Ruckenstein and Schüll 2017). These developments are situated within methodological debates in which social sciences often appear through the perspective of ELSI.

As the field of genomics is progressing rapidly and the intervention in the human genome is no longer science fiction, we argue that it is important to discuss and reflect now on the social configurations that are inscribed in, and reproduced by genomic data-driven research. These may co-produce the conception of certain potentially editable conditions, i.e. create new, and reproduce existing classifications that are largely shaped by societal understandings of difference and order. Such definitions could have real consequences—as Thomas and Thomas (1929) remind us—for individuals and societies, and mark what has been described as an epistemic shift in biomedicine from the clinical gaze to the 'molecular gaze' where the processes of "medicalisation and biomedicalisation both legitimate and compel interventions that may produce transformations in individual, familial and other collective identities" (Clarke et al. 2013, p. 23). While Science and Technology Studies (STS) has demonstrated how science and society are co-produced in research (Jasanoff 2004), we want to use the momentum of the current discourse to critically reflect on these developments from three angles: (1) we demonstrate how, in the process of genomics research, societal relations, understandings and categorizations are used and inscribed into social phenomena and outcomes; (2) we explore the ways that the (big) data-driven research is constituted by increasingly moving away from theory and methodological generation of theoretical concepts that foster the understanding of societal contexts and relations (Kitchin 2014a) and (3) using the GWAS case in focus, we show how the assumption of 'free from theory' (Kichin 2014a) in this case does not mean free of choices made, choices which are themselves restricted by data that are available. We highlight Griffiths' (2016) contention that the material nature of genes, their impacts on biological makeup of individuals and their socially and culturally situated behaviour are not deterministic, and need to be understood within the dynamic, culturally and temporally situated context within which knowledge claims are made. We conclude by making the important point that ignoring the social may lead to a distorted, datafied, genomised body which ignores the key fact that "genes are not stable but essentially malleable" (Prainsack 2015) and that this 'malleability' is rooted in the complex interplay between biological and social environments.

From this perspective, the body is understood through the lens of embodiment, considering humans 'live' their genome within their own lifeworld contexts (Rehmann-Sutter and Mahr 2016). We also consider this paper as an intervention into the marginalization of social science in the wake of developments in data-driven research that neglect social theory, established methodology and the contextual relevance of the social environment.

In the following reflections, we proceed step by step: First, we introduce the case of the GWAS on same-sex sexual behaviour, as well as its limits, context and impact. Second, we recall key sociological theory on categorizations and their implications. Third, we discuss the emergence of a digital-datafication of scientific knowledge production. Finally, we conclude by cautioning against the marginalization of social science in the wake of

developments in data-driven research that neglect social theory, established methodology and the contextual relevance of the social environment.

Studying sexual orientation: The case of same-sex sexual behaviour

Currently, a number of studies at the intersection of genetic and social conditions appear on the horizon. Just as in the examples we have already mentioned, such as those on educational attainment (Lee et al. 2018), or social stratification (Abdellaoui et al. 2019), it is important to note that the limit to such studies is only the availability of the data itself. In other words, once the data is available, there is always the potential that it would eventually be used. This said, an analysis of the entirety of the genomic research on social outcomes and behaviour is beyond the scope of this article. Therefore, we want to exemplify our argument with reference to the research on the genetics of same-sex sexual behaviour.

Based on a sample of half a million individuals of European ancestry, the first large-scale GWAS of its kind claims five genetic variants to be contributing to the assessed "same-sex sexual behaviour" (Ganna et al. 2019b). Among these variants, two are useful only for male–male sexual behaviour, one for female–female sexual behaviour, and the remaining two for both. The data that has led to this analysis was sourced from biobanks/cohorts with different methods of data collection. The authors conclude that these genetic variations are not predictive of sexual orientation; not only because genetics is supposedly only part of the picture, but also because the variations are only a small part (<1% of the variance in same-sex sexual behaviour, p. 4) of the approximated genetic basis (8–25% of the variance in same-sex sexual behaviour) that may be identified with large sample sizes (p. 1). The study is an example of how the 'gay gene' discourse that has been around for years, gets transformed with the available data accumulating in the biobanks and the consequent genomic analysis, offering only one facet of a complex social phenomenon: same-sex sexual behaviour.

The way the GWAS has been conducted was not novel in terms of data collection. Genome-wide studies of similar scale, e.g. on insomnia (Jansen et al. 2019) or blood pressure (Evangelou et al. 2018), often rely on already collected data in biobanks rather than trying to collect hundreds of thousands of individuals' DNA from scratch. Furthermore, in line with wider developments, the study was preregistered[2] with an analysis plan for the data to be used by the researchers. Unlike other GWASes, however, the researchers partnered with an LGBTQIA+ advocacy group (GLAAD) and a science communication charity (Sense About Science), where individuals beyond the research team interpreted the findings and discussed how to convey the results[3]. Following these engagements, the researchers have produced a website[4] with potential frequently asked questions as well as a video about the study, highlighting what it does and what it does not claim.

Despite efforts to control the drifting away of the study into genetic deterministic and discriminatory interpretations, the study has been criticized by many[5]. Indeed, the controversial "How gay are you?"[6] app on the GenePlaza website utilized the findings of the study, which in turn raised the alarm bells and, ultimately, was taken down after much debate. The application, however, showed how rapidly such findings can translate into individualized systems of categorization, and consequently feed into and be fed by the public imaginary. One of the study authors demands continuation of research by noting "[s]cientists have a responsibility to describe the human condition in a more nuanced and deeper way" (Maxmen, 2019, p. 610). Critics, however, note that the context of data collected from the individuals may have

influence on the findings; for instance, past developments (i.e. decriminalization of homosexuality, the HIV/AIDS epidemic, and legalization of same-sex marriage) are relevant to understand the UK Biobank's donor profile and if the GWAS were to be redone according to the birth year of the individuals, different findings could have come out of the study (Richardson et al. 2019, p. 1461).

It has been pointed out that such research should be assessed by a competent ethical review board according to its potential risks and benefits (Maxmen 2019, p. 610), in addition to the review and approval by the UK Biobank Access Sub-Committee (Ganna et al. 2019a, p. 1461). Another ethical issue of concern raised by critics is that the informed consent form of UK Biobank does not specify that it could be used for such research since "homosexuality has long been removed from disease classifications" and that the broad consent forms allow only "health-related research" (Holm and Ploug 2019, p. 1460). We do not want to make a statement here for or against broad consent. However, we argue that discussions about informed consent showcase the complexities related to secondary use of data in research. Similarly, the 'gay gene' app developed in the wake of the sexual orientation study, revealed the difficulty of controlling how the produced knowledge may be used, including in ways that are openly denounced by the study authors.

To the best of our knowledge, there have not been similar genome-wide studies published on sexual orientation and, while we acknowledge the limitations associated with focusing on a single case in our discussion, we see this case as relevant to opening up the following question: How are certain social categorizations incorporated into the knowledge production practices? We want to answer this by first revisiting some of the fundamental sociological perspectives into categorizations and the social implications these may have.

Categorizing sex, gender, bodies, disease and knowledge

Sociological perspectives on categorizations. Categorizations and classifications take a central role in the sociology of knowledge, social stratifications and data-based knowledge production. Categories like gender, race, sexuality and class (and their intersection, see Crenshaw 1989) have become key classifications for the study of societies and in understanding the reproduction of social order. One of the most influential theories about the intertwining of categories like gender and class with power relations was formulated by Bourdieu (2010, 2001). He claimed that belonging to a certain class or gender is an embodied practice that ensures the reproduction of social structure which is shaped by power relations. The position of subjects within this structure reflects the acquired cultural capital, such as education. Incorporated dispositions, schemes of perception, appreciation, classification that make up the individual's *habitus* are shaped by social structure, which actors reproduce in practices. One key mechanism of social categorization is gender classification. The gender order appears to be in the 'nature of things' of biologically different bodies, whereas it is in fact an incorporated social construction that reflects and constitutes power relations. Bourdieu's theory links the function of structuring classifications with embodied knowledge and demonstrates that categories of understanding are pervaded by societal power relations.

In a similar vein Foucault (2003, 2005) describes the intertwining of ordering classifications, bodies and power in his study of the clinic. Understandings of and knowledge about the body follow a specific way of looking at it—the 'medical gaze' of separating the patient's body from identity and distinguishing healthy from the diseased, which, too, is a process pervaded by power differentials. Such classifications evolved historically.

Foucault reminds us that all periods in history are characterized by specific epistemological assumptions that shape discourses and manifest in modalities of order that made certain kinds of knowledge, for instance scientific knowledge, possible. The unnoticed "order of things", as well as the social order, is implemented in classifications. Such categorizations also evolved historically for the discourse about sexuality, or, in particular as he pointed out writing in the late 1970s, distinguishing sexuality of married couples from other forms, such as homosexuality (Foucault 1998).

Bourdieu and Foucault offer two influential approaches within the wider field of sociology of knowledge that provide a theoretical framework on how categorizations and classifications structure the world in conjunction with social practice and power relations. Their work demonstrates that such structuration is never free from theory, i.e. they are not existing prediscursively, but are embedded within a certain temporal and spatial context that constitutes 'situated knowledge' (Haraway 1988). Consequently, classifications create (social) order that cannot be understood as 'naturally' given but as a result of relational social dynamics embedded in power differentials.

Feminist theory in the 1970s emphasized the inherently social dimension of male and female embodiment, which distinguished between biological sex and socially rooted gender. This distinction built the basis for a variety of approaches that examined gender as a social phenomenon, as something that is (re-) constructed in social interaction, impacted by collectively held beliefs and normative expectations. Consequently, the difference between men and women was no longer simply understood as a given biological fact, but as something that is, also, a result of socialization and relational exchanges within social contexts (see, e.g., Connell 2005; Lorber 1994). Belonging to a gender or sex is a complex practice of attribution, assignment, identification and, consequently, classification (Kessler and McKenna 1978). The influential concept of 'doing gender' emphasized that not only the gender, but also the assignment of sex is based on socially agreed-upon biological classification criteria, that form the basis of placing a person in a *sex category*, which needs to be practically sustained in everyday life. The analytical distinction between sex and gender became eventually implausible as it obscures the process in which the body itself is subject to social forces (West and Zimmerman 1991).

In a similar way, sexual behaviour and sexuality are also shaped by society, as societal expectations influence sexual attraction—in many societies within normative boundaries of gender binary and heteronormativity (Butler 1990). This also had consequences for a deviation from this norm, resulting for example in the medicalisation of homosexuality (Foucault 1998).

Reference to our illustrative case study on the recently published research into the genetic basis of sexuality brings the relevance of this theorization into focus. The study cautions against the 'gay gene' discourse, the use of the findings for prediction, and genetic determinism of sexual orientation, noting "the richness and diversity of human sexuality" and stressing that the results do not "make any conclusive statements about the degree to which 'nature' and 'nurture' influence sexual preference" (Ganna et al. 2019b, p. 6).

Coming back to categorizations, more recent approaches from STS are also based on the assumption that classifications are a "spatio-temporal segmentation of the world" (Bowker and Star 2000, p. 10), and that classification systems are, similar to concepts of gender theory (e.g. Garfinkel 1967), consistent, mutually exclusive and complete. The "International Classification of Diseases (ICD)", a classification scheme of diseases based on their statistical significance, is an example of such a historically grown knowledge system. How the ICD is utilized in practice points to the ethical and social dimensions involved (Bowker and

Star 2000). Such approaches help to unravel current epistemological shifts in medical research and intervention, including removal of homosexuality from the disease classification half a century ago.

Re-classifying diseases in tandem with genetic conditions creates new forms of 'genetic responsibilities (Novas and Rose 2000). For instance, this may result in a change of the 'sick role' (described early in Parsons 1951) in creating new obligations not only for diseased but also for actually healthy persons in relation to potential futures. Such genetic knowledge is increasingly produced using large-scale genomic databases and creates new categories based on genetic risk, and consequently, may result in new categories of individuals that are 'genetically at risk' (Novas and Rose 2000). The question now is how these new categories will alter, structure or replace evolved categories, in terms of constructing the social world and medical practice.

While advancement in genomics is changing understandings of bodies and diseases, the meanings of certain social categories for medical research remain rather stable. Developments of personalized medicine go along with "the 're-inscription' of traditional epidemiological categories into people's DNA" and adherence to "old population categories while working out new taxonomies of individual difference" (Prainsack 2015, pp. 28–29). This, again, highlights the fact that knowledge production draws on and is shaped by categories that have a political and cultural meaning within a social world that is pervaded by power relations.

From categorization to social implication and intervention. While categorizations are inherently social in their production, their use in knowledge production has wide ranging implications. Such is the case of how geneticisation of sexual orientation has been an issue that troubled and comforted the LGBTQIA+ communities. Despite the inexistence of an identified gene, 'gay gene' has been part of societal discourse. Such circulation disseminates an unequal emphasis on the biologized interpretations of sexual orientation, which may be portrayed differently in media and appeal to groups of opposing views in contrasting ways (Conrad and Markens 2001). Geneticisation, especially through media, moves sexual orientation to an oppositional framework between individual choice and biological consequence (Fausto-Sterling 2007) and there have been mixed opinions within LGBTQIA+ communities, whether this would resolve the moralization of sexual orientation or be a move back into its medicalisation (Nelkin and Lindee 2004). Thus, while some activists support geneticisation, others resist it and work against the potential medicalisation of homosexuality (Shostak et al. 2008). The ease of communicating to the general public simple genetic basis for complex social outcomes which are genetically more complex than reported, contributes to the geneticisation process, while the scientific failures of replicating 'genetic basis' claims do not get reported (Conrad 1999). In other words, while finding a genetic basis becomes entrenched as an idea in the public imaginary, research showing the opposite does not get an equal share in the media and societal discourse, neither of course does the social sciences' critique of knowledge production that has been discussed for decades.

A widely, and often quantitatively, studied aspect of geneticisation of sexual orientation is how this plays out in the broader understanding of sexual orientation in society. While there are claims that geneticisation of sexual orientation can result in depoliticization of the identities (O'Riordan 2012), it may at the same time lead to polarization of society. According to social psychologists, genetic attributions to conditions are likely to lead to perceptions of immutability, specificity in aetiology, homogeneity and discreteness as well naturalistic fallacy (Dar-Nimrod

and Heine 2011). Despite the multitude of suggestive surveys that belief in genetic basis of homosexuality correlates with acceptance, some studies suggest learning about genetic attribution to homosexuality can be polarizing and confirmatory of the previously held negative or positive attitudes (Boysen and Vogel 2007; Mitchell and Dezarn 2014). Such conclusions can be taken as a precaution that just as scientific knowledge production is social, its consequences are, too.

Looking beyond the case. We want to exemplify this argument by taking a detour to another case where the intersection between scientific practice, knowledge production and the social environment is of particular interest. While we have discussed the social implications of geneticisation with a focus on sexual orientation, recent developments in biomedical sciences and biotechnology also have the potential to reframe the old debates in entirely different ways. For instance, while 'designer babies' were only an imaginary concept until recently, the facility and affordability of processes, such as in vitro selection of baby's genotype and germline genome editing, have potentially important impacts in this regard. When CRISPR/Cas9 technique was developed for rapid and easy gene editing, both the hopes and worries associated with its use were high. Martin and others (2020, pp. 237–238) claim gene editing is causing both disruption within the postgenomic regime, specifically to its norms and practices, and the convergence of various biotechnologies such as sequencing and editing. Against this background, He Jiankui's announcement in November 2018 through YouTube[7] that twins were born with edited genomes was an unwelcome surprise for many. This unexpected move may have hijacked the discussions on ethical, legal, societal implications of human germline genome-editing, but also rang the alarm bells across the globe for similar "rogue" scientists planning experimentation with the human germline (Morrison and de Saille 2019). The facility to conduct germline editing is, logically, only one step away from 'correcting' and if there is a correction, then that would mean a return to a normative state. He's construction of HIV infection as a genetic risk can be read as a placeholder for numerous questions to human germline editing: What are the variations that are "valuable" enough for a change in germline? For instance, there are plans by Denis Rebrikov in Russia to genome edit embryos to 'fix' a mutation that causes congenital deafness (Cyranoski 2019). If legalized, what would be the limits applied and who would be able to afford such techniques? At a time when genomics research into human sociality is booming, would the currently produced knowledge in this field and others translate into 'corrective' genome-editing? Who would decide?

The science, in itself is still unclear at this stage as, for many complex conditions, using gene editing to change one allele to another is often minuscule in effect, considering that numerous alleles altogether may affect phenotypes, while at the same time a single allele may affect multiple phenotypes. In another GWAS case, social genomicists claim there are thousands of variations that are found to be influential for a particular social outcome such as educational attainment (Lee et al. 2018), with each having minimal effect. It has also been shown in the last few years, as the same study is conducted with ever more larger samples, more genomic variants are associated with the social outcome, i.e. 74 single nucleotide polymorphisms (SNPs) associated with the outcome in a sample size of 293,723 (Okbay et al. 2016) and 1271 SNPs associated with the outcome in a sample size of 1.1 million individuals (Lee et al. 2018).

Applying this reasoning to the GWAS on same-sex sexual behaviour, it is highly probable that the findings will be superseded in the following years with similar studies of bigger data, increasing the number of associations.

A genomic re-thinking? The examples outlined here have served to show how focusing the discussion on "genetic determinism" is fruitless considering the complexity of the knowledge production practices and how the produced knowledge could both mirror social dynamics and shape these further. Genomic rethinking of the social necessitates a new formulation of social equality, where genomes are *also* relevant. Within the work of social genomics researchers, there has been cautious optimism toward the contribution of findings from genomics research to understanding social outcomes of policy change (Conley and Fletcher 2018; Lehrer and Ding 2019). Two fundamental thoughts govern this thinking. First, genetic basis is not to be equalized with fate; in other words, 'genetic predispositions' make sense only within the broader social and physical environmental frame, which often allows room for intervention. Second, genetics often relates to heterogeneity of the individuals within a population, in ways that the same policy may be positive, neutral or negative for different individuals due to their genes. In this respect, knowledge gained via social genomics may be imagined as a basis for a more equal society in 'uncovering' invisible variables, while, paradoxically, it may also be a justification for exclusion of certain groups. For example, a case that has initially raised the possibility that policies affect individuals differently because of their genetic background was a genetic variant that was correlated to being unaffected by tax increases on tobacco (Fletcher 2012). The study suggested that raising the taxes may be an ineffective tool for lowering smoking rates below a certain level, since those who are continuing to smoke may be those who cannot easily stop due to their genetic predisposition to smoking. Similar ideas could also apply to a diverse array of knowledge produced in social genomics, where the policies may be under scrutiny according to how they are claimed to variably influence the members of a society due to their genetics.

Datafication of scientific knowledge production

From theory to data-driven science. More than a decade has gone by since Savage and Burrows (2007) described a crisis in empirical research, where the well-developed methodologies for collecting data about the social world would become marginal as such data are being increasingly generated and collected as a by-product of daily virtual transactions. Today, sociological research faces a widely datafied world, where (big) data analytics are profoundly changing the paradigm of knowledge production, as Facebook, Twitter, Google and others produce large amounts of socially relevant data. A similar phenomenon is taking place through opportunities that public and private biobanks, such as UK Biobank or 23andMe, offer. Crossing the boundaries of social sciences and biological sciences is facilitated through mapping correlations between genomic data, and data on social behaviour or outcomes.

This shift from theory to data-driven science misleadingly implies a purely inductive knowledge production, neglecting the fact that data is not produced free of preceding theoretical framing, methodological decisions, technological conditions and the interpretation of correlations—i.e. an assemblage situated within a specific place, time, political regime and cultural context (Kitchin 2014a). It glosses over the fact that data cannot simply be treated as *raw* materials, but rather as "inherently partial, selective and representative", the collection of which has consequences (Kitchin 2014b, p. 3). How knowledge of the body is generated starts with how data is produced and how it is used and mobilized. Through sequencing, biological samples are translated into digital data that are circulated and merged and correlated with other data. With the translation from genes into data, their

meaning also changes (Saukko 2017). The kind of knowledge that is produced is also not free of scientific and societal concepts.

Individually assigned categorical variables to genomes have become important for genomic research and are impacting the ways in which identities are conceptualized under (social) genomic conditions. These characteristics include those of social identity, such as gender, ethnicity, educational and socioeconomic status. They are often used for the study of human genetic variation and individual differences with the aim to advance personalized medicine and based on demographic and ascribed social characteristics.

The sexual orientation study that is central to this paper can be read as a case where such categories intersect with the mode of knowledge production. As the largest contributor of data to the study, UK Biobank's data used in this research are revealing since they are based on the answer to the following question "Have you ever had sexual intercourse with someone of the same sex?" along with the statement "Sexual intercourse includes vaginal, oral or anal intercourse."[8].

Furthermore, the authors accept having made numerous reductive assumptions and that their study has methodological limitations. For instance, Ganna et al. (2019b) acknowledge both within the article (p. 1) and an accompanying website[9] that the research is based on a binary 'sex' system with exclusions of non-complying groups as the authors report that they "dropped individuals from [the] study whose biological sex and self-identified sex/gender did not match" (p. 2). However, both categorizing sexual orientation mainly on practice rather than attraction or desire, and building it on normative assumptions about sexuality, i.e. gender binary and heteronormativity, are problematic, as sexual behaviour is diverse and does not necessarily correspond with such assumptions.

The variations found in the sexual orientation study, as is true for other genome-wide association studies, are often relevant for the populations studied and in this case, those mainly belong to certain age groups and European ancestry. While the study avoids critique in saying that their research is not genetics of sexual orientation, but rather of same-sex sexual behaviour, whether such a genomic study would be possible is also questionable. This example demonstrates that, despite the increasing influence of big data, a fundamental problem with the datafication of many social phenomena is whether or not they are amenable to measurement. In the case of sexual orientation, whether the answer to the sexual orientation questions corresponds to the "homosexuality" or "willingness to reveal homosexuality"/"stated sexual orientation" is debatable, considering the social pressure and stigma that may be an element in certain social contexts (Conley 2009, p. 242).

While our aim is to bring a social scientific perspective, biologists have raised at least two different critical opinions on the knowledge production practice here in the case of the sexual orientation study, first on the implications of the produced knowledge[10] and second on the problems and flaws of the search for a genetic basis[11]. In STS, however, genetic differences that were hypothesized to be relevant for health, especially under the category of *race* in the US, have been a major point of discussion within the genomic 'inclusion' debates of 1990s (Reardon 2017, p. 49; Bliss 2015). In other words, a point of criticism towards the knowledge production was the focus on certain "racial" or racialized groups, such as American of European ancestry, which supposedly biased the findings and downstream development of therapies for 'other' groups. However, measuring health and medical conditions against the background of groups that are constituted based on social or cultural categories (e.g. age, gender, ethnicity), may also result in a reinscription/reconstitution of social inequalities attached to these categories (Prainsack 2015) and at the same time result in health justice being a topic seen

through a postgenomics lens, where postgenomics is "a frontline weapon against inequality" (Bliss 2015 p. 175). Social-economic factors may recede in the background, while data with its own often invisible politics are foregrounded.

Unlike what Savage and Burrows suggested in 2007, the coming crisis can not only be seen as a crisis of sociology, but of science in general. Just as the shift of focus in social sciences towards digital data is only one part of the picture, another part could be the developments in genomisation of the social. Considering that censuses and large-scale statistics are not new, the distinction of the current phenomenon is possibly the opportunity to individualize the data, while categories themselves are often unable to capture the complexity, despite producing knowledge more efficiently. In that sense, the above-mentioned survey questions do not do justice to the complexity of social behaviour. What is most important to flag within these transformations is the lack of reflexivity regarding how big data comes to represent the world and whether it adds and/or takes away from the ways of knowing before big data. These developments and directions of genetic-based research and big data go far beyond the struggle of a discipline, namely sociology, with a paradigm shift in empirical research. They could set the stage for real consequences for individuals and groups. Just as what is defined as an editable condition happens as a social process that relies on socio-political categories, the knowledge acquired from big data relies in similar way on the same kind of categories.

The data choices and restrictions: 'Free from theory' or freedom of choice. Data, broadly understood, have become a fundamental part of our lives, from accepting and granting different kinds of consent for our data to travel on the internet, to gaining the 'right to be forgotten' in certain countries, as well as being able to retrieve collected information about ourselves from states, websites, even supermarket chains. While becoming part of our lives, the data collected about individuals in the form of big data is transferred between academic and non-academic research, scientific and commercial enterprises. The associated changes in the knowledge production have important consequences for the ways in which we understand and live in the world (Jasanoff 2004). The co-productionist perspective in this sense does not relate to whether or how the social and the biological are co-produced, but rather it is pointing to how produced knowledge in science is both shaped by and shaping societies. Thus, the increasing impact and authority of big data in general, and within the sexual orientation study in focus here, opens up new avenues to claim as some suggest, that we have reached the end of theory.

The "end of theory" has actively been debated within and beyond science. Kitchin (2014a) locates the recent origin of this debate in a piece in the *Wired*, where the author states "Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all" (Anderson 2008). Others call this a paradigm shift towards data-intensive research leaving behind the empirical and theoretical stages (Gray 2009, p. xviii). While Google and others form the basis for this data-driven understanding in their predictive capacity or letting the data speak, the idea that knowledge production is 'free from theory' in this case seems to be, at best, an ignorance of any data infrastructure and how the categories are formed within it.

Taking a deeper look at the same-sex sexual behaviour study from this angle suggests that such research cannot be free from theory as it has to make an assumption regarding the role of genetics in the context of social dynamics. In other words, it has to move sexual orientation, at least partially in the form of same-sex sexual behaviour, out of the domain of the social towards the

biological. In doing so, just as the study concludes the complexity of sexual orientation, the authors note in their informative video[12] on their website, that "they found that about a third of the differences between people in their sexual behaviour could be explained by inherited genetic factors. But the environment also plays a large role in shaping these differences." While the study points to a minuscule component of the biological, it also frames biology as the basis on which the social, as part of the environment, acts upon.

Reconsidering how the biology and the social are represented in the study, three theoretical choices are made due to the limitation of the data. First of all, the biological is taken to be "the genome-wide data" in the biobanks that the study relies on. This means sexual orientation is assumed to be within the SNPs, or points on the genome that are common variations across a population, and not in other kinds of variations that are rare or not captured by the genotyped SNPs. These differences include, but are not limited to, large-scale to small-scale duplications and deletions of the genomic regions, rare variants or even common variants in the population that the SNP chips do not capture. Such ignored differences are very important for a number of conditions, from cancer to neurobiology. Similarly, the genomic focus leaves aside the epigenetic factors that could theoretically be the missing link between genomes and environments. In noting this, we do not suggest that the authors of the study are unaware or uninterested in epigenetics; however, regardless of their interest and/or knowledge, the availability of large-scale genome-wide data puts such data ahead of any other variation in the genome and epigenome. In other words, if the UK Biobank and 23andMe had similar amounts of epigenomic or whole genome data beyond the SNPs, the study would have most possibly relied on these other variations in the genome. The search for genetic basis within SNPs is a theoretical choice, and in this case this choice is pre-determined by the limitations of the data infrastructures.

The second choice that the authors make is to take three survey questions, i.e. in the case of UK Biobank data, as encompassing enough of the complexity of sexual orientation for their research. As partly discussed earlier, these questions are simply asking about sexual behaviour. Based on the UK Biobank's definition of sexual intercourse as "vaginal, oral or anal intercourse" the answers to the following questions were relevant for the research: "Have you ever had sexual intercourse with someone of the same sex?" (Data-Field 2159), "How many sexual partners of the same sex have you had in your lifetime?" (Data-Field 3669), and, "About how many sexual partners have you had in your lifetime?" (Data-Field 2149). Answers to such questions do little justice to the complexity of the topic. Considering that they are not included in the biobank as data for the purpose of identifying a genetic basis to same-sex sexual behaviour, there is much to consider in what capacity they are useful for that. It is worth noting here that the UK Biobank is primarily focused on health-related research, and thus these three survey questions could not have been asked with a genomic exploration of 'same-sex sexual behaviour' or 'sexual orientation' in mind. The degree of success in the way they have been used to identify the genetic basis for complex social behaviours is questionable.

The authors of the study consider the UK Biobank sample to be comprised of relatively old individuals and this to be a short-coming[13]. Similarly, the study authors claim that 23andMe samples may be biased because "[i]ndividuals who engage in same-sex sexual behaviour may be more likely to self-select the sexual orientation survey", which then explains the high percentage of such individuals (18.9%) (Ganna et al. 2019b, p. 1). However, the authors do not problematize that there is at least three-fold difference between the youngest and oldest generation in the UK Biobank sample in their response to the same-sex sexual behaviour

question (Ganna et al. 2019b, p. 2). The study, thus, highlights the problematic issue about who should be regarded as the representative sample to be asked about their "same-sex sexual behaviour". Still, this is a data choice that the authors make in concluding a universal explanation out of a very specific and socially constrained collection of self-reported data that encompasses only part of what the researchers are interested in.

The third choice is a choice unmade. The study data mainly came from UK Biobank, following a proposal by Brendan Zietsch with the title "Direct test whether genetic factors predisposing to homosexuality increase mating success in heterosexuals"[14]. The original plan for research frames "homosexuality" as a condition that heterosexuals can be "predisposed" to and as this condition is not eliminated through evolution, scientists hypothesize that whatever genetic variation that predisposes an individual to homosexuality may also be functional in increasing the individual's reproductive capacity. Despite using such an evolutionary explanation as the theoretical basis for obtaining the data from the UK Biobank, the authors use evolution/evolutionary only three times in the article, whereas the concept "mating success" is totally missing. Unlike the expectation in the research proposal, authors observe lower number of offspring for individuals reporting same-sex sexual behaviour, and they conclude briefly "This reproductive deficit raises questions about the evolutionary maintenance of the trait, but we do not address these here" (Ganna et al. 2019b, p. 2). In other words, the hypothesis that allowed scientists to acquire the UK Biobank data becomes irrelevant for the researchers, when they are reporting their findings.

In this section, we have performed an analysis of how data choices are made at different steps of the research and hinted at how these choices reflect certain understandings of how society functions. These are evident in the ways sexual behaviour is represented and categorized according to quantitative data, and, the considerations of whether certain samples are contemporary enough (UK Biobank) or too self-selecting (same-sex sexual behaviour being too high in 23andMe). The study, however, does not problematize how the percentage of individuals reporting same-sex sexual behaviour steadily increases according to year of birth, at least tripling for males and increasing more than five-fold for females from 1940 and 1970 (for UK Biobank). Such details are among the data that the authors display as descriptive statistics in Fig. 1 (Ganna et al. 2019b, p. 2); however, these do not attract a discussion that genomic data receives. The study itself starts from the idea that genetic markers that are associated with same-sex sexual behaviour could have an evolutionary advantage and ends in saying the behaviour is complex. Critics claim the "approach [of the study] implies that it is acceptable to issue claims of genetic drivers of behaviours and then lay the burden of proof on social scientists to perform post-hoc socio-cultural analysis" (Richardson et al. 2019, p. 1461).

Conclusion

In this paper, we have 'moved back to the future'—taking stock of the present-day accelerated impact of big data and of its potential and real consequences. Using the sexual orientation GWAS as point of reference, we have shown that claims to working under the premise of 'pure science' of genomics are untenable as the social is present by default—within the methodological choices made by the researchers, the impact on/of the social imaginary or epigenetic context.

By focusing on the contingency of the knowledge production on the social categories that are themselves reflections of the social in the data practices, we have highlighted the relational processes at the root of knowledge production. We are experiencing a period where the repertoire of what gets quantified continuously, and possibly exponentially, increases; however, this

does not necessarily mean that our understanding of complexity increases at the same rate, rather, it may lead to unintended simplification where meaningful levels of understanding of causality are lost in the "triumph of correlations" in big data (Mayer-Schönberger and Cukier 2013; cited in Leonelli 2014). While sociology has much to offer through its qualitative roots, we think it should do more than critique, especially considering the culturally and temporally specific understandings of the social are also linked to the socio-material consequences.

We want to highlight that now is the time to think about the broader developments in science and society, not merely from an external perspective, but within a new framework. Clearly, our discussion of a single case here cannot sustain suggestions for a comprehensive and applicable framework for any study; however, we can flag the urgency of its requirement. We have shown that, in the context of the rapid developments within big data-driven, and socio-genomic research, it is necessary to renew the argument for bringing the social, and its interrelatedness to the biological, clearly back into focus. We strongly believe that reemphasizing this argument is essential to underline the analytical strength of the social science perspective, and in order to avoid the possibility of losing sight of the complexity of social phenomena, which risk being oversimplified in mainly statistical data-driven science.

We can also identify three interrelated dimensions of scientific practice that the framework would valorize: (1) Recognition of the contingency of choices made within the research process, and sensibility of their consequent impact within the social context. (2) Ethical responsibilities that move beyond procedural contractual requirements, to sustaining a process rooted in clear understanding of societal environments. (3) Interdisciplinarity in analytical practice that potentiates the impact of each perspectival lens.

Such a framework would facilitate moving out of the disciplinary or institutionalized silos of ELSI, STS, sociology, genetics, or even emerging social genomics. Rather than competing for authority on 'the social', the aim should be to critically complement each other and refract the produced knowledge with a multiplicity of lenses. Zooming 'back to the future' within the field of socio-biomedical science, we would flag the necessity of re-calibrating to a multi-perspectival endeavour—one that does justice to the complex interplay of social and biological processes within which knowledge is produced.

Notes

1 The GWAS primarily uses the term "same-sex sexual behaviour" as one of the facets of "sexual orientation" where the former becomes the component that is directly associable with the genes and the latter the broader phenomenon of interest. Thus, while the article is referring to "same-sex sexual behaviour" in its title, it is editorially presented in the same *Science* issue under Human Genetics heading with the subheading "The genetics of sexual orientation" (p. 880) (see Funk 2019). Furthermore, the request for data from UK Biobank by the corresponding author Brendan P. Zietsch (see footnote 14) refers only to sexual orientation and homosexuality and not to same-sex sexual behaviour. Therefore, we follow the same interchangeable use in this article.
2 Source: https://osf.io/xwfe8 (04.03.2020).
3 Source: https://www.wsj.com/articles/research-finds-genetic-links-to-same-sex-behavior-11567101661 (04.03.2020).
4 Source: https://geneticsexbehavior.info (04.03.2020).
5 In addition to footnotes 10 and 11, for a discussion please see: https://www.nytimes.com/2019/08/29/science/gay-gene-sex.html (04.03.2020).
6 Later "122 Shades of Grey": https://www.geneplaza.com/app-store/72/preview (04.03.2020).
7 Source: https://www.youtube.com/watch?v=th0vnOmFltc (04.03.2020).

8 Source: http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2159 (04.03.2020).
9 Source: https://geneticsexbehavior.info/ (04.03.2020).
10 Source: https://www.broadinstitute.org/blog/opinion-big-data-scientists-must-be-ethicists-too (04.03.2020).
11 Source: https://medium.com/@cecilejanssens/study-finds-no-gay-gene-was-there-one-to-find-ce5321c87005 (03.03.2020).
12 Source: https://videos.files.wordpress.com/2AVNyj7B/gosb_subt-4_dvd.mp4 (04.03.2020).
13 Source: https://geneticsexbehavior.info/what-we-found/ (04.03.2020).
14 Source: https://www.ukbiobank.ac.uk/2017/04/direct-test-whether-genetic-factors-predisposing-to-homosexuality-increase-mating-success-in-heterosexuals/ (04.03.2020).

References

Abdellaoui A, Hugh-Jones D, Yengo L, Kemper KE, Nivard MG, Veul L, Holtz Y, Zietsch BP, Frayling TM, Wray NR (2019) Genetic correlates of social stratification in Great Britain. Nat Hum Behav 1–21. https://doi.org/10.1038/s41562-019-0757-5

Anderson C (2008) The end of theory: the data deluge makes the scientific method obsolete, Wired https://www.wired.com/2008/06/pb-theory/. Accessed 31 Mar 2020

Bliss C (2015) Defining health justice in the postgenomic era. In: Richardson SS, Stevens H (eds) Postgenomics: perspectives on biology after the genome. Duke University Press, Durham, Durham/London, pp. 174–191

Bourdieu P (2001) Masculine domination. Stanford University Press, Stanford

Bourdieu P (2010) Distinction: a social critique of the judgement of taste. Routledge, London/New York

Bowker GC, Star SL (2000) Sorting things out: classification and its consequences. MIT Press, Cambridge/London

Boysen GA, Vogel DL (2007) Biased assimilation and attitude polarization in response to learning about biological explanations of homosexuality. Sex Roles 57(9–10):755–762. https://doi.org/10.1007/s11199-007-9256-7

Butler J (1990) Gender trouble. Feminism and the subversion of identity. Routledge, New York

Clarke AE, Shim JK, Shostak S, Nelson A (2013) Biomedicalising genetic health, diseases and identities. In: Atkinson P, Glasner P, Lock M (eds) Handbook of genetics and society: mapping the new genomc era. Routledge, Oxon, pp. 21–40

Conley D (2009) The promise and challenges of incorporating genetic data into longitudinal social science surveys and research. Biodemogr Soc Biol 55 (2):238–251. https://doi.org/10.1080/19485560903415807

Conley D, Fletcher J (2018) The genome factor: what the social genomics revolution reveals about ourselves, our history, and the future. Princeton University Press, Princeton/Oxford

Connell RW (2005) Masculinities. Polity, Cambridge

Conrad P (1999) A mirage of genes. Sociol Health Illn 21(2):228–241. https://doi.org/10.1111/1467-9566.00151

Conrad P, Markens S (2001) Constructing the 'gay gene' in the news: optimism and skepticism in the US and British press. Health 5(3):373–400. https://doi.org/10.1177/136345930100500306

Crenshaw K (1989) Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics, vol 1989(8). University of Chicago Legal Forum. http://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8. Accessed 1 Apr 2020

Cyranoski D (2019) Russian 'CRISPR-baby' scientist has started editing genes in human eggs with goal of altering deaf gene. Nature 574(7779):465–466. https://doi.org/10.1038/d41586-019-03018-0

Dar-Nimrod I, Heine SJ (2011) Genetic essentialism: on the deceptive determinism of DNA. Psychol Bull 137(5):800–818. https://doi.org/10.1037/a0021860

Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, Ntritsos G, Dimou N, Cabrera CP, Karaman I (2018) Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat Genet 50(10):1412–1425. https://doi.org/10.1038/s41588-018-0205-x

Fausto-Sterling A (2007) Frameworks of desire. Daedalus 136(2):47–57. https://doi.org/10.1162/daed.2007.136.2.47

Fletcher JM (2012) Why have tobacco control policies stalled? Using genetic moderation to examine policy impacts. PLoS ONE 7(12):e50576. https://doi.org/10.1371/journal.pone.0050576

Foucault M (1998) The history of sexuality 1: the will to knowledge. Penguin Books, London

Foucault M (2003) The birth of the clinic. Routledge, London/New York

Foucault M (2005) The order of things. Routledge, London/New York

Fox Keller E (2014) From gene action to reactive genomes. J Physiol 592 (11):2423–2429. https://doi.org/10.1113/jphysiol.2014.270991

Fox Keller E (2015) The postgenomic genome. In: Richardson SS, Stevens H (eds) Postgenomics: perspectives on biology after the genome. Duke University Press, Durham/London, pp. 9–31

Funk M (2019) The genetics of sexual orientation. Science 365(6456):878–880. https://doi.org/10.1126/science.365.6456.878-k

Ganna A, Verweij KJ, Nivard MG, Maier R, Wedow R, Busch AS, Abdellaoui A, Guo S, Sathirapongsasuti JF, Lichtenstein P (2019a) Genome studies must account for history—response. Science 366(6472):1461–1462. https://doi.org/10.1126/science.aaz8941

Ganna A, Verweij KJ, Nivard MG, Maier R, Wedow R, Busch AS, Abdellaoui A, Guo S, Sathirapongsasuti JF, Lichtenstein P (2019b) Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior. Science 365(6456):eaat7693. https://doi.org/10.1126/science.aat7693

Garfinkel H (1967) Studies in ethnomethodology. Polity Press, Cambridge

Gray J (2009) Jim Gray on eScience: a transformed scientific method. In: Hey T, Tansley S, Tolle KM (eds) The fourth paradigm: data-intensive scientific discovery. Microsoft Research, Redmond, pp. xvii–xxxi

Griffiths DA (2016) Queer genes: realism, sexuality and science. J Crit Realism 15 (5):511–529. https://doi.org/10.1080/14767430.2016.1210872

Hamer DH, Hu S, Magnuson VL, Hu N, Pattatucci AM (1993) A linkage between DNA markers on the X chromosome and male sexual orientation. Science 261(5119):321–327. https://doi.org/10.1126/science.8332896

Haraway D (1988) Situated knowledges: the science question in feminism and the privilege of partial perspective. Fem Stud 14(3):575–599

Holm S, Ploug T (2019) Genome studies reveal flaws in broad consent. Science 366 (6472):1460–1461. https://doi.org/10.1126/science.aaz3797

Howard HC, van El CG, Forzano F, Radojkovic D, Rial-Sebbag E, de Wert G, Borry P, Cornel MC (2018) One small edit for humans, one giant edit for humankind? Points and questions to consider for a responsible way forward for gene editing in humans. Eur J Hum Genet 26(1):1. https://doi.org/10.1038/s41431-017-0024-z

Jansen PR, Watanabe K, Stringer S, Skene N, Bryois J, Hammerschlag AR, de Leeuw CA, Benjamins JS, Muñoz-Manchado AB, Nagel M, Savage JE, Tiemeier H, White T, Agee M, Alipanahi B, Auton A, Bell RK, Bryc K, Elson SL, Fontanillas P, Furlotte NA, Hinds DA, Huber KE, Kleinman A, Litterman NK, McCreight JC, McIntyre MH, Mountain JL, Noblin ES, Northover CAM, Pitts SJ, Sathirapongsasuti JF, Sazonova OV, Shelton JF, Shringarpure S, Tian C, Wilson CH, Tung JY, Hinds DA, Vacic V, Wang X, Sullivan PF, van der Sluis S, Polderman TJC, Smit AB, Hjerling-Leffler J, Van Someren EJW, Posthuma D, The 23andMe Research, T. (2019) Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nat Genet 51(3):394–403. https://doi.org/10.1038/s41588-018-0333-3

Jasanoff S (2004) The idiom of co-production. In: Jasanoff S (ed.) States of knowledge: the co-production of science and social order. Routledge, London, p 1–12

Jasanoff S, Hurlbut JB (2018) A global observatory for gene editing. Nature 555:435–437. https://doi.org/10.1038/d41586-018-03270-w

Kessler SJ, McKenna W (1978) Gender: an ethnomethodological approach. John Wiley & Sons, New York

Kitchin, R. (2014a) Big Data, new epistemologies and paradigm shifts. Big Data Soc. https://doi.org/10.1177/2053951714528481

Kitchin R (2014b) The data revolution. Big data, open data, data infrastructures and their consequences. Sage, London

Landecker H (2016) The social as signal in the body of chromatin. Sociol Rev 64 (1_suppl):79–99. https://doi.org/10.1111/2059-7932.12014

Landecker H, Panofsky A (2013) From social structure to gene regulation, and back: a critical introduction to environmental epigenetics for sociology. Annu Rev Sociol 39:333–357. https://doi.org/10.1146/annurev-soc-071312-145707

Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, Nguyen-Viet TA, Bowers P, Sidorenko J, Linnér RK (2018) Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. Nat Genet 50(8):1112. https://doi.org/10.1038/s41588-018-0147-3

Lehrer SF, Ding W (2019) Can social scientists use molecular genetic data to explain individual differences and inform public policy? In: Foster G (ed.) Biophysical measurement in experimental social science research. Academic Press, London/San Diego/Cambridge/Oxford, pp. 225–265

Leonelli, S. (2014) What difference does quantity make? On the epistemology of Big Data in biology. Big Data Soc. https://doi.org/10.1177/2053951714534395

Lorber J (1994) Paradoxes of gender. Yale University Press, New Haven

Martin P, Morrison M, Turkmendag I, Nerlich B, McMahon A, de Saille S, Bartlett A (2020) Genome editing: the dynamics of continuity, convergence, and change in the engineering of life. New Genet Soc 39(2):219–242. https://doi.org/10.1080/14636778.2020.1730166

Maxmen A (2019) Controversial 'gay gene' app provokes fears of a genetic Wild West. Nature 574(7780):609–610. https://doi.org/10.1038/d41586-019-03282-0

Mayer-Schönberger V, Cukier K (2013) Big data: a revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt, Boston/New York

Meloni M (2014) Biology without biologism: social theory in a postgenomic age. Sociology 48(4):731–746. https://doi.org/10.1177/0038038513501944

Meloni M (2016) Political biology: Science and social values in human heredity from eugenics to epigenetics. Palgrave Macmillan, n.p.p

Mitchell RW, Dezarn L (2014) Does knowing why someone is gay influence tolerance? Genetic, environmental, choice, and "reparative" explanations. Sex Cult 18(4):994–1009. https://doi.org/10.1007/s12119-014-9233-6

Morrison M, de Saille S (2019) CRISPR in context: towards a socially responsible debate on embryo editing. Palgrave Commun 5(1):1–9. https://doi.org/10.1057/s41599-019-0319-5

Nelkin D, Lindee MS (2004) The DNA mystique: the gene as a cultural icon. University of Michigan Press, Ann Arbor

Novas C, Rose N (2000) Genetic risk and the birth of the somatic individual. Econ Soc 29(4):485–513. https://doi.org/10.1080/03085140050174750

O'Riordan K (2012) The life of the gay gene: from hypothetical genetic marker to social reality. J Sex Res 49(4):362–368. https://doi.org/10.1080/00224499.2012.663420

Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, Turley P, Chen G-B, Emilsson V, Meddens SFW (2016) Genome-wide association study identifies 74 loci associated with educational attainment. Nature 533 (7604):539–542. https://doi.org/10.1038/nature17671

Parry B, Greenhough B (2017) Bioinformation. Polity Press, Cambridge

Parsons T (1951) The social system. Free Press, New York

Prainsack B (2015) Is personalized medicine different? (Reinscription: the sequel) A response to Troy Duster. Br J Sociol 66(1):28–35. https://doi.org/10.1111/1468-4446.12117

Reardon J (2017) The postgenomic condition: ethics, justice, and knowledge after the genome. University of Chicago Press, Chicago/London

Rehmann-Sutter C, Mahr D (2016) The lived genome. In: Whitehead A, Woods A (eds) Edinburgh companion to the critical medical humanities. Edinburgh University Press, Edinburgh, pp. 87–103

Richardson SS, Borsa A, Boulicault M, Galka J, Ghosh N, Gompers A, Noll NE, Perret M, Reiches MW, Sandoval JCB (2019) Genome studies must account for history. Science 366(6472):1461. https://doi.org/10.1126/science.aaz6594

Ruckenstein M, Schüll ND (2017) The datafication of health. Annu Rev Anthropol 46(261–278). https://doi.org/10.1146/annurev-anthro-102116-041244

Saukko P (2017) Shifting metaphors in direct-to-consumer genetic testing: from genes as information to genes as big data. New Genet Soc 36(3):296–313. https://doi.org/10.1080/14636778.2017.1354691

Savage M, Burrows R (2007) The coming crisis of empirical sociology. Sociology 41 (5):885–899. https://doi.org/10.1177/0038038507080443

Shostak S, Conrad P, Horwitz AV (2008) Sequencing and its consequences: path dependence and the relationships between genetics and medicalization. Am J Sociol 114(S1):S287–S316. https://doi.org/10.1086/595570

Thomas WJ, Thomas DS (1929) The child in America. Behavior problems and programs. Knopf, New York

West C, Zimmerman DH (1991) Doing gender. In: Lorber J, Farrell SA (eds) The social construction of gender. Sage, Newbury Park/London, pp. 13–37

Acknowledgements

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.G. or K.A.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.