Blue Chips and White Collars: Whose Data Science Is It?

Seliem El-Sayed^{1,2} Barbara Prainsack³

¹Governance of Digital Practices Research Platform, University of Vienna, Vienna, Austria,

²Centre for the Study of Contemporary Solidarity, Department of Political Science, University of Vienna, Vienna, Austria,

³Governance of Digital Practices Research Platform, University of Vienna, Vienna, Austria,
Department of Political Science, University of Vienna, Vienna, Austria

Published on: Jan 29, 2021

DOI: <https://doi.org/10.1162/99608f92.ea29ee5a>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

One of the most important parts of [Leonelli's article](#), we believe, is her insightful analysis of the epistemic and ontological assumptions inherent in practices of data use. Leonelli argues that three assumptions underpinning data use are particularly problematic as they support the belief in a technological fix: First, that data are “reliable and unambiguous in the information they convey”; second, that data can be easily transformed into decision making in various policy fields, and third, that data do not have any harmful consequences for democratic governance in the longer term (Leonelli, 2021). A few months after Leonelli's text was written, her analysis reads as a prescient explanation of what policymakers have done wrong in their attempts to constrain the COVID-19 crisis over much of the summer and autumn of 2020.

In this short commentary, we want to underscore some points in Leonelli's analysis and proposal that we believe are particularly important for data science in the next phase of the pandemic, and also expand on some elements that we believe require further discussion. We will end with the proposal that the translation of findings from data science into policy interventions is not only, as Leonelli states, the task of politicians, but scientists as well. But first, let us turn to the challenge that is, in our view, the most urgent to address: The inclusion of experiential and contextual knowledge into data science.

“You Cannot ‘Big Data’ Your Way Out of a ‘No Data’ Situation”: Capturing and Communicating Experiential and Contextual Knowledge for Pandemic Decision Making

Leonelli draws attention to the fact that not all data are created equal: some data, particularly those that can be ‘cleaned and fitted into available models,’ are prioritized over others. They are favored not only because of the flawed yet tenacious assumption that quantified and computable data collected are superior to qualitative data that still need to be packaged into countable and computable categories. In some cases, it is also a matter of time constraints: faster is better. But the idea that qualitative data is a ‘nice to have’ once the essentials have been covered needs to change. If we leave unaddressed the problem that certain kinds of questions can only be answered on the basis of experiential and contextual data (that may not be ready computation, or may not have been captured even), then we will remain unable to develop policy measures that are effective. In fact, we rely on experiential knowledge already at the stage of problem-framing if we wish to avoid epistemic injustice—a point we return to in the following.

To give an example from our own part of the world: following the announcements that COVID-19 vaccines may be close to approval in Europe, the Austrian government started to devise plans for deployment and priority setting. The government acknowledged uncertainties regarding long-term side effects of the different vaccines, the extent to which they would prevent a serious disease course, as well as protect from infecting others. In addition, policymakers kept an attentive eye on public opinion surveys; these indicated that people's willingness to get vaccinated against COVID-19 was around 30% (Kittel, 2020; Kittel et al., 2020a). Those

who were skeptical of vaccinations were soon branded by experts as lacking solidarity, and some politicians started calling for vaccination mandates (Löwenstein, 2020). Needless to say that this moral politics did not appease public concerns at all; moreover, rumors about a possible vaccination mandate for health care workers caused concern about a mass exodus of the workforce, from care homes in particular.

The reasons for which many people were skeptical of vaccines were explored in a few studies: For example, data from a qualitative study (that the authors of this commentary were involved in) showed that the dividing line between vaccination skeptics and supporters in the context of COVID-19 did not map neatly onto that between pro- and anti-vaxxers in other contexts (Kieslich et al., 2020). Findings from a representative survey showed that people who believed that their immune system could fend off the SARS-Cov2-virus corresponded strongly with vaccination hesitancy (Kittel et al., 2020a, 2020b, 2020c). Moreover, for some of the respondents in both studies, skepticism of vaccinations against COVID-19 was associated with a general distrust of the government's ability to deal with the crisis. Despite these efforts to collect and analyze experiential data in a rapid manner, and communicating these findings to publics and policy makers, these findings on the deeper reasons for people's behaviors did not influence policy making. Had policy makers recognized the value of such knowledge, it would have been possible to communicate more effectively, and to build the factors that influence vaccination uptake for COVID-19 into national and regional vaccination policies and plans.

Our proposal here is not merely that data on the experiential knowledge of people, and qualitative data on people's practices, should be used to complement the analysis of 'hard' data when findings from the latter seem inconsistent or otherwise in need of further explanation. On the contrary, especially at a time when so many of the key features of our societies—from child care to work to psychological wellbeing and physical health—are different from those in times of normal societal functioning, insights into people's practices and experiences need to be collected and analyzed in systematic ways, and their value for policy making needs to be recognized. We need to know what health care workers, teachers, and home-schooling parents experience, and how they deal with and interpret the situation they are in, so that we can alleviate problems effectively and be better prepared for the next crisis. These are questions that cannot be answered on the basis of data that people capture with and on their smartphones. Experiential knowledge (first coined by Thomasina Borkman, 1976, to refer to the knowledge of patients, but more widely applicable to the knowledge that emerges from personal and professional practice; see Cook & Wagenaar, 2012) is, thus, not only important to be included in the interpretation of data, but it also needs to guide decisions on *what* data we collect.

This would require a shift in the focus of our expert and public discussions on data science. Our policy and societal discourses prominently address how data can be made more accessible, and how barriers to interoperability can be removed (e.g., European Commission, 2020; Organisation for Economic Co-operation and Development [OECD], 2020). But we talk very little about what data we want to collect in the first place, and about how and why it matters. What aspects of our bodies and lives do we datafy, and what do we not? With what consequences?

The Role of Theory Is Even More Important When There Is Pressure to Do Without

In addition to the importance of a more inclusive view on *what* data should be used to inform policies, we also must consider *how* data has been curated and analyzed (Leonelli, 2016). The extent to which theory must precede data collection is contested and varies across different fields. Some social scientists have been skeptical of computational methods used for research in social sciences because computer scientists might be unfamiliar with previous social theory and draw conclusions based on patterns rather than causality (Ledford, 2020).¹ Data collected without knowledge about the context, and data analysis on the basis of simplistic models of the underlying causes, they argue, are of little value. For them, theory must precede if not guide data collection (e.g., Wallach, 2015).

Others, in contrast, have criticized that “many social-science theories are too nebulous to be tested using big data” (Ledford, 2020). Concepts, even if well established in social science research, are not always easily translated into a quantitative format that can be read by computers. How, for instance, could one measure social capital? Rather than social science theory being ‘nebulous,’ we argue, this problem shows that computational models are not appropriate to address just any scientific question. While we do not want to dispute that there is, of course, some poor social theory, theoretical concepts that retain a certain level of openness often do so to be able to accommodate the complexities of the social realities that they capture. If such social theory cannot ‘be tested using big data,’ then this is the case because the data and models are too simplistic to fit the concept. (This is not a critique of models as such; models are helpful *because* they are abstractions, not in spite of it. But, as Boulton et al. (2015, p. 97) remind us, “mathematical models can give these abstractions a misleading solidity” and take the model, rather than the social phenomena that they seek to capture, as a normative reference point.) It would be wrong, thus, to conclude that the solution is to throw out the theory and keep using the data. Instead, we need to expand the range of epistemological and methodological approaches (and, as argued above, the types of data) that we include in the term ‘data science.’

We also share Leonelli’s concern about too casual an attitude to causal inference (see also Hernán & Robins, 2020, which Leonelli cites in her article, this issue). This casual attitude can stem from an erroneous conflation of correlation and causation (such as in Beral et al., 1982, and Boyce, 2018).² But it can also be driven by the assumption that the difference between correlation and causation does not matter. An example for this latter view is Chris Anderson’s now (in)famous *Wired* article advocating for a departure from the scientific method. Anderson proclaimed the “end of theory” and held that in an era where “[c]orrelation is enough.” We can stop looking for models. We can analyze the data without hypotheses about what it might show” (Anderson, 2008; see also Hilbert, 2016). Anderson argues that, in the process of producing scientific knowledge, theory is becoming obsolete and the reasons for which “people do what they do” cease to matter. Although Anderson’s article was heavily criticized, in some ways, the idea that data collection for knowledge production does neither require previous theorization nor (even simplistic) models of causality has gained some traction (Coveney et

al., 2016; Wise & Shaffer, 2015). This is particularly true in the context of the COVID-19 pandemic, where time constraints can be used to justify the need to act before understanding the underlying causal pathways.

Consider this example of hypothesis-free data mining. In a pre-print paper from June 2020, MIT's Christopher R. Knittel and Microsoft's Bora Ozaltun correlated U.S. county-level COVID-19 death rates with a range of socioeconomic variables, county-level health variables, modes of commuting, and climate and pollution patterns. Although emphasizing that their observational analyses only identified correlations that are not “necessarily causal,” they suggest that the results “may help policy makers identify variables that may potentially be causally related to COVID-19 death rates and adopt appropriate policies *after understanding the causal relationship*” [emphasis added] (Knittel & Ozaltun, 2020, p. 5). One rather surprising finding reported in their paper is that “counties with higher home values have higher death-rates” (Baskin, 2020). According to the authors of the study, the reasons for this are unclear.³ Despite the authors' disclaimer that the correlations identified in their study may not represent causal relationships, their findings may still be considered actionable when pressure on policymakers mounts to design policy based on those scientific results. But without additional research, this result is, in essence, uninterpretable.

Seeing the increasing frequency with which the results of theory-free data mining are portrayed as actionable, medical researcher Ayoub and colleagues (2020) recently urged that, since “‘big data’ and ‘real-world data’ movements have gained an increasing foothold in the pages of high-impact journals during the pandemic” it is more important than ever to care about the difference between correlation and causation. Using the “strong positive correlation between a country's FIFA⁴ ranking and their COVID-19 ranking” as an example, they emphasized that it is impossible to decide whether these results are meaningful without recourse to theory (Ayoub et al., 2020). They urged the academic community and policy-makers alike to not “forget that outcomes of observational studies are only hypothesis forming.”

In sum, while hypothesis-free data mining can be a useful activity in the early stages of scientific knowledge creation, we should only act upon its results when we understand the nature of their relationship. In addition, amid a global pandemic, findings yielded by big data analysis that represent society, individual behaviors, or bodies at a stage of ‘normal’ functioning may not always be helpful for a society in crisis. Also in this sense, the COVID-19 pandemic demonstrates the limits of purely correlation-based artificial intelligence models, which have difficulties predicting values for the exceptional situation we find ourselves in and for which there is no data to feed a model with (Benaich, 2020; Steinhardt & Toner, 2020). What are the assumptions that are inherent in the data sets, and in the method of analysis? What ‘baseline’ do the data reflect? Next to questions about the underlying causal pathway, these are questions that need to be addressed before political measures follow the findings of data mining.

Not One Public Benefit

Lastly, we will consider how there is not one public benefit, but how benefits accrue to different publics to different extents. Scholars from different fields, such as science and technology studies (STS), communication studies but also biochemistry, have long emphasized that technology is inherently normative (see, among others, Beckwith, 1986; Christians, 1989; Irwin & Wynne, 2003; Jasanoff, 2004; Rose & Rose 1979). Since the 1980s various writers have suggested normative frameworks by which “technological process [...] can proceed responsibly” (Christians, 1989, p. 123) and benefit the public (Radder, 2019). Radder, for example, proposes assessing whether technology benefits the public by evaluating the “relative quality of democratic support” in electoral, constitutional, and deliberative terms (2019, p. 205). Implicit (or explicit) conceptualizations of how a technology benefits the public differ between specific approaches and underline the ambiguity of the phenomenon of public benefit and interest.

An even more pressing task than figuring out how we could adequately define public interest, at this point in the crisis, is an exploration of how the technological tools represented by, and emerging from, data science benefit people differently—both within and across societies. That technologies (understood in the wider sense of the word) yield different benefits and costs for different groups is, of course, neither unique to data science, nor to COVID-19. Consider the example of electric vehicles, where numerous ‘public benefits’ come to mind: cleaner air will lead to overall improved health in the population and less noisy streets are also in the interest of the public. A closer look, however, reveals that the benefits of electric vehicles disproportionately accrue to wealthy members of society. First, as more and more electric vehicles fill the streets, often generously subsidized by governments (International Energy Agency, 2020), it becomes apparent that the mobility of the future still includes cars. Consequently, governments will continue to invest in the maintenance and development of infrastructure related to the use of cars. People who cannot afford cars, and rely on public transport such as trains, benefit less from the development and subsidy of electric cars than wealthier people (Lesser, 2018).

Thus, even with technologies that are portrayed as ‘clean’ and ‘good,’ benefits accrue differently and can amplify existing inequities related to access to technology and wealth. Policymakers and researchers alike should be wary to attest that a technology benefits ‘the public.’ Rather, how a technology affects different groups in society, who benefits and to what extent, and who may be harmed must be carefully explored in ways that include the voices of all societal groups. This is particularly important in a global crisis where the costs and burdens are so unequally distributed (Bottan et al., 2020; The Health Foundation, 2020).

Doing so is not only a matter of distributional justice in the sense of the equitable allocation of scarce resources (Schmidt, 2020); but it is also a matter of epistemic justice—a fair and equitable way of defining the problems that we are collectively solving, and choosing how to solve them. This ties back into our first point on the need for experiential knowledge. Where we disproportionately prioritize data that can be “cleaned and fitted into

available models” (Leonelli, 2020) we are unlikely to capture any problems expressed in data emanating from experiential knowledge. It is precisely that knowledge, however, that can help us avoid epistemic injustice.

Conclusion: Data Science Cannot Be Apolitical

Professor Leonelli’s analysis (this issue) is both helpful and astute in explaining where we can go wrong, and where we can go right, with our use of data science in times of a global crisis. The one point where we disagree with her is that “translating... scientific findings into interventions remains the responsibility of politicians rather than scientists.” Scientists, we argue, have a role in this too: While they are not to do the work of politicians in terms of finding majorities and managing public budgets, scientists can no longer pretend that they are merely doing neutral analyses either. It is not only, as Leonelli also emphasizes, that data do not speak for themselves, and that the task of interpretation requires the expertise and experiential knowledge of scientists and practitioners alike. The possible influence of scientists on policymaking starts with choosing the questions they prioritize—which is embodied also in Florence Nightingale’s sentence that the “main end of statistics should not be to inform the government as to how many men have died, but to enable immediate steps to be taken to prevent the extension of disease and mortality” (Nightingale, 1958, cited in McDonald, 2012, p. 328; see also, Office for Statistics Regulation, 2020, p. 5).

Scientists also have impact on policy in problem framing. The very way that scientific problems are framed suggest certain solutions. To remain with our earlier example of vaccinations, if scientists frame the problem of vaccination hesitancy as one of insufficient information, then the ‘solution’ will ignore that hesitation in participating in a vaccination program run by public authorities can articulate much more than people’s views on vaccinations: it can express a lack of trust in government, or a generalized fear in light of the very high uncertainties that people live with (see also Paul & Loer, 2019). It would miss the insight that, to reduce people’s uncertainties, economic and social support goes a long way, and that the best antidote to conspiracy theories could be the reduction of economic and social inequalities. In gathering different types of knowledge, including people’s real and lived experiences, researchers can play an important role in translating the insights from data science into political measures. Data science in times of pandemics, thus, also means that ‘translation’ does not only start with publishing our study’s results, but it starts with the translation of an idea into a research design. It continues with organizing data so that they can answer relevant questions, and the inclusion of an ‘ethics of modeling’ in all data science (see King, 2020). Such an ethics should not merely tick the boxes of research ethics in the narrow sense, ensuring that privacy and data protection standards have been met. It is a political ethics in the sense that asks how data science affects the distribution of resources, power, and agency across and within populations.

We are not suggesting that data science should become political. We are arguing that it already is. The way we do and say things in science has an impact on policymaking, whether we like it or not. What the pandemic does, however, is to place a stronger obligation on scientists to explicitly reflect on the norms, goals, and values that our research articulates and promotes.

Acknowledgments

The authors thank Tom King, Elisabeth Steindl, Hendrik Wagenaar and Lukas Schlögl for helpful comments on earlier drafts. All mistakes remain ours.

Disclosure Statement

Seliam El-Sayed and Barbara Prainsack have no financial or non-financial disclosures to share for this article.

References

- Anderson, C. (2008, June 23). The end of theory: The data deluge makes the scientific method obsolete. *Wired*. <https://www.wired.com/2008/06/pb-theory/>
- Ayoub, F., Sato, T., & Sakuraba, A. (2020). Football and COVID-19 risk: Correlation is not causation. *Clinical Microbiology and Infection*, 27(2), 291–292. <https://doi.org/10.1016/j.cmi.2020.08.034>
- Barrowman, N. (2014). Correlation, causation, and confusion. *The New Atlantis* (Summer/Fall). <https://www.thenewatlantis.com/publications/correlation-causation-and-confusion>
- Baskin, K. (2020, July 15). *4 unexpected findings about COVID-19 deaths*. MIT Sloan School of Management. <https://mitsloan.mit.edu/ideas-made-to-matter/4-unexpected-findings-about-covid-19-deaths>
- Beckwith, J. (1986). The radical science movement in the United States. *Monthly Review*, 38(3), 118–129. https://doi.org/10.14452/mr-038-03-1986-07_12
- Benaich, N. (2020, September 20). AI has disappointed on Covid. *Financial Times*. <https://www.ft.com/content/0aafc2de-f46d-4646-acfd-4ed7a7f6feaa>
- Beral, V., Shaw, H., Evans, S., & Milton, G. (1982). Malignant melanoma and exposure to fluorescent lighting at work. *The Lancet*, 320(8293), 290–293. [https://doi.org/10.1016/s0140-6736\(82\)90270-7](https://doi.org/10.1016/s0140-6736(82)90270-7)
- Borkman, T. (1976). Experiential knowledge: A new concept for the analysis of self-help groups. *Social Service Review*, 50(3), 445–456. <https://doi.org/10.1086/643401>
- Bottan, N., Hoffmann, B. & Vera-Cossio, D. (2020). The unequal impact of the coronavirus pandemic: Evidence from seventeen developing countries. *PLOS ONE*, 15(10), Article e0239797. <https://doi.org/10.1371/journal.pone.0239797>
- Boulton, J. G., Allen, P. M. & Bowman, C. (2015). *Embracing complexity: Strategic perspectives for an age of turbulence*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199565252.001.0001>

Boyce, P. (2018). Editorial: Correlation, causation and confusion. *Lighting Research & Technology*, 50(5), 657. <https://doi.org/10.1177/1477153518787733>

Buehlman, K. T., Gottman, J. M. & Katz, L. F. (1992). How a couple views their past predicts their future: Predicting divorce from an oral history interview. *Journal of Family Psychology*, 5(3-4), 295–318. <https://doi.org/10.1037/0893-3200.5.3-4.295>

Christians, C. (1989). A theory of normative technology. In Byrne, E.F. & Pitt, J.C. (Eds.), *Technological Transformation* (pp. 123–139). Springer. https://doi.org/10.1007/978-94-009-2597-7_9

Cook, S. N. & Wagenaar, H. (2012). Navigating the eternally unfolding present: Toward an epistemology of practice. *The American Review of Public Administration*, 42(1), 3–38. <https://doi.org/10.1177/0275074011407404>

Coveney, P. V., Dougherty, E. R. & Highfield, R. R. (2016). Big data need big theory too. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2080), Article 20160153. <https://doi.org/10.1098/rsta.2016.0153>

Donovan, J. [@BostonJoan]. (2019, August 22). Quietly, reading this paper in @NatureNews “Hidden resilience and adaptive dynamics of the global online hate ecology” and it really misses the mark on how hate movements propagate and should be moderated... [Twitter thread]. <https://threader.app/thread/1164542554886553601>

European Commission. (2020). *European health data space*. Retrieved December 28, 2020, from https://ec.europa.eu/health/ehealth/dataspace_en

The Health Foundation. (2020). *The same pandemic, unequal impacts*. Retrieved December 31, 2020, from <https://www.health.org.uk/news-and-comment/charts-and-infographics/same-pandemic-unequal-impacts>

Hernán, M. A., & Robins, J. M. (2020). *Causal inference: What if*. Chapman & Hill/CRC. <https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>

Hilbert, M. (2016). Big data for development: A review of promises and challenges. *Development Policy Review*, 34(1), 135–174. <https://doi.org/10.1111/dpr.12142>

International Energy Agency. (2020). *Global EV outlook 2020*. <https://www.iea.org/reports/global-ev-outlook-2020>

Irwin, A., & Wynne, B. (Eds.) (2003). *Misunderstanding science? The public reconstruction of science and technology*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511563737>

Jasanoff, S. (2004). Ordering knowledge, ordering society. In S. Jasanoff (Ed.), *States of knowledge: The co-production of science and the social order* (pp. 13–45). Routledge. <https://doi.org/10.4324/9780203413845-7>

Kieslich, K., El-Sayed, S., Haddad, C., Paul, K., Pot, M., Prainsack, B., Radhuber, I., Schlögl, L., Spahl, W., & Weiss, E. (2020). From New Forms of Community to Fatigue: How the Discourse about the COVID-19 Crisis Has Changed. *Solidarity in times of a pandemic (SolPan) Blog*. Retrieved December 23, 2020, from https://digigov.univie.ac.at/solidarity-in-times-of-a-pandemic-solpan/solpan-blog-english/blog-posts/news/from-new-forms-of-community-to-fatigue-how-the-discourse-about-the-covid-19-crisis-has-changed-1/?tx_news_pi1%5Bcontroller%5D=News&tx_news_pi1%5Baction%5D=detail&cHash=8a5c34534be62ffe5c578d6721435eeb

King, T. (2020). Ethics in a time of SARS 2. *Public Statistics*. Retrieved December 31, 2020, from <https://publicstatistic.wordpress.com/2020/05/13/ethics-in-a-time-of-sars-2/>

Kittel, B. (2020). Die Erosion der Impfbereitschaft in der österreichischen Bevölkerung (The Erosion of the Willingness to Get Vaccinated in the Austrian Population). *Vienna Center for Electoral Research – Corona Blog*. Retrieved December 23, 2020, from <https://viecer.univie.ac.at/en/projects-and-cooperations/austrian-corona-panel-project/corona-blog/corona-blog-beitraege/blog87/>

Kittel, B., Kritzing, S., Boomgaarden, H., Prainsack, B., Eberl, J.-M., Kalleitner, F., Lebernegg, N., Partheymüller, J., Plescia, C., Schiestl, D. & Schlögl, L. (2020a). The Austrian Corona Panel Project: Monitoring individual and societal dynamics amidst the COVID-19 crisis. *European Political Science*, 20(2), 318–344. <https://doi.org/10.1057/s41304-020-00294-7>

Kittel, B., Kritzing, S., Boomgaarden, H., Prainsack, B., Eberl, J.-M., Kalleitner, F., Lebernegg, N., Partheymüller, J., Plescia, C., Schiestl, D. & Schlögl, L. (2020b). Austrian Corona Panel Project (SUF edition). <https://doi.org/10.11587/28KQNS>

Kittel, B., Paul, K., Kieslich, K. and Resch, T. (2020c). Die Impfbereitschaft Der Österreichischen Bevölkerung Im Dezember 2020 (The Willingness Of The Austrian Population To Get Vaccinated In December 2020). *Vienna Center for Electoral Research – Corona Blog*. Retrieved January 22, 2021, from <https://viecer.univie.ac.at/corona-blog/corona-blog-beitraege/blog93>

Knittel, C. R. & Ozaltun, B. (2020). What does and does not correlate with COVID-19 death rates. *medRxiv*. <https://doi.org/10.1101/2020.06.09.20126805>

Ledford, H. (2020, June 17). How Facebook, Twitter and other data troves are revolutionizing social science. *Nature*. <https://www.nature.com/articles/d41586-020-01747-1>

- Leonelli, S. (2016). *Data-centric biology: A philosophical study*. The University of Chicago Press.
<https://doi.org/10.7208/chicago/9780226416502.001.0001>
- Leonelli, S. (2021). Data science in times of Pan(dem)ic. *Harvard Data Science Review*, 3(1).
<https://doi.org/10.1162/99608f92.fbb1bdd>
- Lesser, J. (2018, May 15). Are electric cars worse for the environment? *Politico*.
<https://www.politico.com/agenda/story/2018/05/15/are-electric-cars-worse-for-the-environment-000660/>
- Löwenstein, S. (2020). Wenn sich kaum einer testen lässt: Österreich diskutiert nach Massentest über Impfpflicht (When hardly anyone gets tested: Austria discusses obligatory vaccinations after mass test). *Frankfurter Allgemeine Zeitung*. Retrieved December 28, 2020, from
<https://www.faz.net/aktuell/politik/ausland/oesterreich-diskutiert-nach-massentest-ueber-impfpflicht-17103904.html>
- McDonald, L. (Ed.). (2012). *Florence Nightingale and hospital reform: Collected works of Florence Nightingale* (Vol. 16). Wilfrid Laurier University Press.
- Organisation for Economic Co-operation and Development. (2020). Why open science is critical to combatting COVID-19. Retrieved December 20, 2020, from <http://www.oecd.org/coronavirus/policy-responses/why-open-science-is-critical-to-combatting-covid-19-cd6ab2f9/>
- Office for Statistics Regulation. (2020). *The public good of statistics: What we know so far*. Retrieved December 30, 2020, from <https://osr.statisticsauthority.gov.uk/publication/the-public-good-of-statistics-what-we-know-so-far/>
- Paul, K. T., & Loer, K. (2019). Contemporary vaccination policy in the European Union: Tensions and dilemmas. *Journal of Public Health Policy*, 40(2), 166–179. <https://doi.org/10.1057/s41271-019-00163-8>
- Radder, H. (2019). *From commodification to the common good: Reconstructing science, technology, and society*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctvnb7qrh>
- Rose, H., & Rose, S. (1979). Radical science and its enemies. *Socialist Register*, 16, 317–335.
<https://socialistregister.com/index.php/srv/article/view/5443/2342>
- Schmidt, H. (2020, April 15). The way we ration ventilators is biased. *The New York Times*.
<https://www.nytimes.com/2020/04/15/opinion/covid-ventilator-rationing-blacks.html>
- Steinhardt, J., & Toner, H. (2020, June 8). Why robustness is key to deploying AI. *Brookings*.
<https://www.brookings.edu/techstream/why-robustness-is-key-to-deploying-ai/>

Wallach, H. (2015). *Computational social science* [Video]. 32nd International Conference on Machine Learning, Lille 2015. Videolectures.net. http://videolectures.net/icml2015_wallach_social_science/

Wise, A. F., & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics*, 2(2), 5–13. <https://doi.org/10.18608/jla.2015.22.2>

©2021 Seliem El-Sayed and Barbara Prainsack. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](https://creativecommons.org/licenses/by/4.0/), except where otherwise indicated with respect to particular material included in the article.

Footnotes

1. Joan Donovan gives the example of computer scientists who authored a study published in *Nature* mapping a network of online hate groups. Because the authors were unfamiliar with previous research on the topic, they could not make adequate sense of their results (see Donovan, 2019). [↵](#)
2. See also Barrowman (2014) on how newspapers and magazines can confuse correlation with causation in their portrayal of research results of Buehlmann et al. (1992). [↵](#)
3. Knittel speculates that “higher home values [might be] correlated with denser homes, living in cities, and so on. So that [point] might be picking up the fact that people just live closer to each other” (Baskin, 2020). [↵](#)
4. Fédération Internationale de Football Association. [↵](#)