Epistemic Injustice and Algorithmic Epistemic Injustice in Healthcare

Jeffrey Byrnes, *Grand Valley State University* (*United States)*
Andrew Spear, Grand Valley State University (United States)
International Conference on Computer Ethics: Philosophical Enquiry (CEPE) 2023, Chicago, IL

Keywords: epistemic injustice, medical epistemic injustice, algorithmic injustice, automation bias

Extended Abstract

We argue that the introduction of algorithmic support systems into medical decision-making, while holding out much promise, also exacerbates ethical concerns deriving from existing knowledge- and power-asymmetries between healthcare providers on the one hand and patients on the other. In several areas, issues with which medicine is already struggling threaten to become more ethically fraught as algorithmic systems enter the picture. Worse still, the very authority accorded to such systems might serve to cover over or render invisible this fraughtness.

Healthcare involves a delicate balance between medical expertise and skill on the one hand and respect for the personhood, autonomy, and insights of patients on the other. While medical expertise and skill are indispensable, properly diagnosing and treating patients requires taking their experiences, insights about their own condition, and health outcome goals seriously. Striking this balance is challenging because there is an asymmetry of knowledge and power that privileges the authority of healthcare professionals. At the same time, patients find themselves in disorienting and vulnerable conditions, and for which they typically lack the technical training and vocabulary necessary to advocate for themselves. While this situation raises well-recognized ethical issues concerning duties of beneficence and nonmaleficence, and patient autonomy, it also has a distinctly epistemological component: medical training and expertise confer epistemic authority for a number of reasons, but this does not mean that patient testimony, input, or preferences have nothing distinctive to contribute to the full medical understanding of their condition.

Building on these insights, Kidd and Carel (2018) have recently argued that the healthcare context is one where epistemic injustice occurs and has some likelihood to occur. According to Fricker (2007), *epistemic injustice* occurs when an individual is wronged specifically in their capacity as a knower: as a reasoning, judging, and believing being. The two main types of epistemic injustice, introduced by Fricker and discussed in subsequent literature, are testimonial injustice and hermeneutic injustice. Testimonial injustice occurs "…when negative stereotyping leads a hearer to prejudicially deflate the credibility assigned to a speaker" (Kidd and Carel, p. 215), while hermeneutical injustice occurs when "…the capacity of a person or group to make intelligible certain of their bodily, existential, and social experiences to themselves or to others is unjustly constrained or undermined" (p. 219). An example of testimonial injustice would be if a female's reporting of what happened or contribution to a discussion were not given the credibility it merited due to negative prejudice concerning the sincerity or competence of women on the part of her hearer(s). In such a case she is harmed "as a knower" by not being permitted to contribute to shared knowledge and understanding of the situation on an equal footing with others. An example of hermeneutical injustice might be police harassment of an African American (say during a traffic stop or a stop and search) just because the person is an African American at a time when the idea or concept of *racial profiling* was not yet well-developed in our cultural vocabulary and legal system. Such an individual suffers both the injustice of being stopped due to prejudice, and the further

distinctively epistemic injustice of not being able to fully conceptualize and communicate to others via shared concepts the wrong that has been done to them.

In Kidd and Carel's analysis, the healthcare context is particularly ripe for epistemic injustice due not to bad actors, but rather due to the understanding of health and disease that currently dominate Western medicine. More specifically, it is the prevalence of objective, naturalistic conceptions of health and disease, such as those expressed in the work of Christopher Boorse (1975 and onward), which define health and the goals of healthcare entirely in objective scientific terms that patients do not and typically cannot be expected to fully understand, that ground healthcare professionals in the (largely inadvertent) epistemic injustices they commit. Testimonial injustice in the medical context involves unwarranted downgrading of the credibility of a patient's testimony and requests on grounds that those who are ill are ignorant, incapable of clear thinking, or otherwise dominated by their condition. Hermeneutical injustice is likely to occur either as a result of the specialized terminologies and protocols of medicine of which patients are typically ignorant or as a result of being part of a patient group whose particular conditions or modes of presenting have been understudied in systematic ways (e.g., a possible consequence of the longstanding practice of treating male anatomy as canonical in research and trials).

We extend the framework of Fricker, Kidd, and Carel to address the consequences of introducing big data and machine-learning-driven algorithmic classification and decision-support systems into healthcare for epistemic injustice in healthcare. We argue that the introduction of sophisticated algorithmic systems and machine learning into healthcare—for example, systems that make recommendations concerning further treatment, medication, or hospitalization— mirrors and threatens to further exacerbate the epistemic injustices discussed by Kidd and Carel.

Concerning testimonial injustice, what little credibility patients may have been taken to have now threatens to be further eroded. In addition to the, at times excessive, trust already placed in the epistemic authority of healthcare professionals, patients will now confront the additional authority typically afforded to automated systems (so-called "automation bias"). If this "two against one" scenario was not bad enough, there are legitimate concerns that current systems of this sort encode biases from the broader medical and social context, thus potentially importing other kinds of testimonial injustice (toward certain groups) into the healthcare context.

Concerning hermeneutical injustice, patients must now contend not only with (to them) esoteric medical concepts and bureaucracy, but also with often opaque and inscrutable deliverances of algorithm-driven systems. Even where the opacity of AI systems of this sort is due primarily to the level of technical understanding needed to interpret them, the operations of such systems are still likely to be opaque to healthcare providers and patients alike, creating new hermeneutical barriers for patients in the understanding of their own conditions that may not be surmountable even with the full assistance and cooperation of trained medical professionals themselves.

Finally, if, as Kidd and Carel maintain, part of what grounds epistemic injustice in healthcare is over-reliance on a particularly objective and scientific conception of health, it seems clear that the same attitude is only further entrenched and supported by reliance on the supposedly objective and scientific deliverances of algorithmic systems.

Our argument is not a wholesale one against the deployment of machine learning algorithms and systems based on them in medicine as such. In particular, it is likely that in relatively specific areas such as the analysis of test results or images, such systems are or soon will be highly successful. Even here, caution is required so that existing tendencies to epistemic injustice are not further exacerbated or reinforced. However, our concern is most centrally with the deployment, especially the overly confident deployment, of such systems in the context of sensitive and often fraught synthetic judgments about the nature, course, and continuation of patient care. In such cases, we argue that the tendency for epistemic injustice to occur is already quite high, and that the deployment of algorithmic systems is likely to make it still higher.

Part of the promise of algorithmic systems is that they might help render certain decisions more consistent and more efficient under conditions of high stress and limited resources. Yet for this very reason there is a significant danger that their mere deployment will be taken to have solved the problem (under conditions of limited resources and capacity, who will have the time to circle back and check anyway?) and may well lead to a devaluing or abdication of the human capacities typically needed to make such judgments. If the need for humanities-informed ethical decision-making in medicine has already been driven home in recent decades by changes in medical treatment, practices, and technology, then the addition of algorithmic systems to medicine does not alleviate, but rather further exacerbates this need. We thus call for a renewed emphasis on training and continuing education in the human dimensions and judgments involved in medical care as a countermeasure.

References

1. Boorse C. (2002). A Rebuttal on functions. In *Functions: New essays in the philosophy and psychology of biology*. Edited by Ariew A, Cummins R, Perlman M. Oxford: Oxford University Press, 63–112.
2. Boorse, C., 1977. "Health as a Theoretical Concept," *Philosophy of Science*, 44: 542– 573.
3. Boorse, C. (1975) "On The Distinction Between Disease and Illness," *Philosophy and Public Affairs*, 5: 49–68.
4. Burrell, J. (2016). *How the machine 'thinks': Understanding opacity in machine learning algorithms*. Big Data & Society Jan—June, 1—12.
5. Floridi, L. (2017). Infra Ethics—On the Conditions of Possibility of Morality. Philos. Technol. (2017) 30:391–394
6. Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford: Oxford University Press.
7. Kearns, M., & Roth, A. (2020). *The ethical algorithm: The science of socially aware algorithm design*. New York, NY: Oxford University Press.
8. Kidd, I. J., & Carel, H. (2017). Epistemic injustice and illness. *Journal of Applied Philosophy, 34*(2), 172-190.
9. Carel, H. & Kidd, I. J. (2014). "Epistemic injustice in healthcare: A philosophical analysis". *Med Health Care and Philos* 17: 529-540.
10. Carel, H. (2008/2019). *Illness: The cry of the flesh*. London: Routledge.
11. Ledford, H. (2019) Millions affected by racial bias in health-care algorithm. *Nature* 574, 608 –

609. Accessed February 10, 2020: https://www.nature.com/magazine assets/d41586-019-03228-6/d41586-019-03228-6.pdf
12. Lyell, D., & Coiera, E. (2017;2016;). Automation bias and verification complexity: A systematic review. *Journal of the American Medical Informatics Association, 24*(2), 423- 431. doi:10.1093/jamia/ocw105
13. Lyell, D., Magrabi, F., Raban, M. Z., Pont, L. G., Baysari, M. T., Day, R. O., & Coiera, E. (2017). Automation bias in electronic prescribing. *BMC Medical Informatics and Decision Making, 17*(1), 28. doi:10.1186/s12911-017-0425-5
14. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 1—21.
15. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science (New York, N.Y.), 366*(6464), 447-453. doi:10.1126/science.aax2342
16. Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future — big data, machine learning, and clinical medicine. *The New England Journal of Medicine, 375*(13), 1216- 1219. doi:10.1056/NEJMp1606181
17. Watson, D. et al. (2019). Clinical applications of machine learning algorithms: beyond the black box. BMJ 364, 1886.