MATHEMATICAL VALUES AND THE EPISTEMOLOGY OF DATA PRACTICES

PATRICK ALLO

Contemporary data practices, whether we call them data science or AI, statistical learning or machine learning, are widely perceived to be game changers. They change what is at stake epistemologically as well as ethically. This especially applies to decision-making processes that infer new insights from data, use these insights to decide on the most beneficial action, and refer to data and to an inference process to justify the chosen course of action. The profiling of citizens is now only one of many such processes.

One of the original goals of 'Profiling the European Citizen' was to understand the nature of the knowledge that data-mining creates and that profiles encode, and to critically assess the epistemic power that is exerted when a profile is applied to an individual. When we develop a critical epistemology for contemporary data practices, we still seek answers to the same questions. We want to know what kind of knowledge is being created, how we may evaluate it, and how it acquires its epistemic authority.

Developing a critical epistemology that does not merely restate the promises of data-driven inquiry, but instead allows us to understand the threats it may pose is a non-trivial task. There is a lack of clarity regarding the epistemological norms we should adhere to. Purely formal evaluations of decisions under uncertainty can, for instance, be hard to assess outside of the formalism they rely on. In addition, there is substantial uncertainty with regard to the applicable norms because scientific norms may appear to be in flux (new paradigms, new epistemologies, etc.) Finally, dealing with this uncertainty and lack of clarity is further complicated by promises of unprecedented progress and opportunities that invite us to imagine a data-revolution with many guaranteed benefits, but few risks.

My goal in this provocation is to focus on a small, easily disregarded, fragment of this broader epistemological project. The inquiry I would like to propose questions the role of mathematics and the role of our beliefs about the nature of mathematical knowledge within contemporary data-practices. What I contend is that, first, there are few reasons to leave the role of mathematics unexamined, and, second, that a conscious reflection on how mathematical thought shapes contemporary data-practices is a fruitful new line of inquiry. It forces us to look beyond data and code (the usual suspects of the critical research agenda on data) and can help us grasp how the epistemic authority of data science is construed.

The role of mathematics

Mathematics does not only contribute to the theoretical foundations of many existing data practices (from sheer counting to learning, categorising, and predicting), but it also contributes to the scientific respectability and trustworthiness of data science. Reliance on mathematics does not only enable (no calculation without mathematics) and certify (no correct calculation without mathematics) data science, but it also makes it credible. Following one of the central motivations of the Strong Programme in the Sociology of Science, I take the task of 'explain[ing] the credibility of a given body of knowledge in given context' (Barnes 1982, xi) to be essential for understanding the epistemology of data science. We should direct our attention to the epistemic authority of mathematics, the epistemic authority granted by mathematics to its applications, and the view that relying on mathematics is epistemically as well as ethically commendable.

An analysis of the role of mathematics in data science that seeks to account for the credibility and authority of data science can be fruitfully developed with an explicit reference to mathematical values. This can help us understand the epistemological contribution of mathematics to data science. It reveals how mathematics, for many the one source of absolute certainty we have, could have any substantial influence on the epistemology of fallible or merely probable predictions. Certainty and truth, of course, are mathematical values, but so is the importance that is accorded to abstract reasoning, or the requirement that the only acceptable proofs and calculations are those that can independently be verified. By attending to such values, we can discern more clearly the influence of mathematical thought within the realm of uncertain reasoning. This is a first advantage of conceiving of the role of mathematics in terms of the values it promotes and the values it appeals to. In addition, when we shift our attention to values we are no longer restricted to a strict accuracy-centric assessment of probabilistic procedures. The latter perspective is traditionally associated with a consequentialist understanding of good decisions. Instead, we can follow a more flexible assessment that lets us to address additional socio-epistemic requirements like trust, responsibility, or accountability.

The critical evaluation of the role of mathematics in data science should not be reduced to the uncovering of the crushing power of the authority of mathematics, or the dismissal of the mathematically warranted neutrality of algorithmic processes. Instead, we should strive to re-think the ambivalent role of mathematics and of beliefs about mathematics in data science. The interaction between mathematics and data science is bi-directional. Data science appeals to mathematical values—such as objectivity, neutrality, and universality—to legitimate itself, but mathematics also promotes certain values—such as the openness of mathematical justification through proof and calculation—in the knowledge practices that rely on mathematics. I contend that data science seeks to associate itself to mathematical values it fails to live up to, but also that some of some mathematical values are not necessarily virtuous when deployed outside the realm of pure mathematics. Mathematical values can be used critically, for instance by underscoring the epistemic value of practically verifiable calculations, but they can also be used in less critical ways, for instance when mathematical techniques are presented as value-free technological artefacts.

A detailed overview of mathematical values is beyond the scope of the present contribution (I refer the interested reader to the seminal contributions of Alan Bishop and Paul Ernest on whose work I draw, e.g. Bishop 1991; Ernest 2016). I will now just focus on one value to illustrate the ambivalent influence of mathematical values on the epistemology of data science. I propose to focus on the importance that mathematical practices accord to 'closed texture' and will argue that as a property of concepts that is closely associated with the demands of abstraction, precision, and explicitness in mathematical reasoning, it is a perfect example of a janus-faced value that can have beneficial as well as detrimental consequences in contexts where mathematical techniques are used to derive actionable knowledge from messy data.

Open and closed texture

The notion of 'open texture' was first coined by Friedrich Waismann (1945) to refer to the fact that many concepts or words we use to describe the world are such that the linguistic rules that govern their use do not determinately settle all their possible uses. Some use-cases appear to be open or unlegislated:

> *The fact that in many cases there is no such thing as a conclusive verification is connected to the fact that most of our empirical concepts are not delimited in all possible directions. (…) Open texture, then, is something like possibility of vagueness. Vagueness can be remedied by giving more accurate rules, open texture cannot (Waissman quoted in Shapiro 2006, 210–1).*[1]

Closed texture, then, is the absence of open texture. Mathematics and computing crucially depend on the absence of open texture, where the absence of unlegislated cases is associated with such values as clarity, explicitness, and univocality. The relevance of the contrast between open and closed texture is based on the paradoxical situation that, on the one hand, the semi-technical notion of an algorithm, understood as a procedure that can be executed without having to rely on the ingenuity or informed judgement of the executor of that procedure, is built on the assumption of closed texture, whereas, on the other hand, the concepts we use to deal with the world (so-called empirical concepts) exhibit open texture. Colours in the world exhibit open texture, but the values of a pixel do not; similarly, the properties of a data-subject may be underdetermined, but the values we find in each field of a data-base are, again, a determinate manner. It is because data, or 'capta' (Kitchin and Dodge 2011), especially when understood as simple syntactical objects, do not exhibit open texture that they are fit for algorithmic processing. This rudimentary insight is easily forgotten when learning-algorithms are deployed for tasks, like image-recognition, for which our human ability to interpret and use concepts that exhibit open texture or are imprecise cannot be captured in precise rules. Whether a given image shows a cat is arguably not something that can be mechanically decided, but whether a collection of pixels does or does not match a given pattern can be so decided. The goal of a learning algorithm is precisely to find a good enough replacement of problems of the former type with problems of the latter type.

This much should be uncontroversial but does not yet explain why 'closed texture' is a janus-faced requirement of mathematical reasoning and of algorithmic processing. This requires us to see that while (as I have just argued) closed texture is a technical requirement of any computational process, its epistemological import is not unequivocally positive. This is because, whereas aspiring to clarify as well as one can the concepts one uses is naturally perceived as an intellectual virtue and as a way to avoid fallacies of equivocation, the closed texture of our concepts is often no more than a convenient (but false) assumption.

Proxies and their target

Let me, to conclude this provocation, briefly describe the risks that are associated with the assumption that all our algorithms operate in the absence of open texture. The risk in question is that the technical need to avoid open texture is easily turned into what van Deemter (2010) calls 'false clarity': our tendency to use imprecise concepts as if they were crisp. Because we replace a question of interest ('is this a cat?') that may not have a determinate answer with a proxy-problem that does have a determinate answer ('is this pattern present?') and can therefore be algorithmically resolved, it is tempting to confuse our ability to correctly solve the proxy-problem with our ability to provide a correct answer to the actual problem. This is especially problematic when the (mathematically supported) trust we place in the former is directly transferred to the latter. It is even more so when a question on which we can reasonably disagree (or whose resolution is context-dependent) is replaced by a question that can be resolved in a controlled environment that does not admit disagreement. In such cases, adherence to the demands of algorithmic processes may spill over into the unwarranted dismissal of critical objections because we confuse the impossibility of disagreeing about the (mathematically represented) proxy-problem with the possibility of disagreeing about the (real-world) target-problem.

Notes

[1] My exposition builds on Shapiro (2006), which focuses more directly on the role of the open and closed texture of concepts within the formal sciences than Hart's seminal work on the open texture of legal rules (Hart & Green 2012; Schauer 2013).

References

Barnes, Barry. 1982. T. S. Kuhn and Social Science. London and Basingstoke: MackMillan.

Bishop, Alan J. 1991. Mathematical Enculturation: A Cultural Perspective on Mathematics Education. Dordrecht, Boston: Kluwer Academic Publishers.

Ernest, Paul. 2016. "Mathematics and Values." In Mathematical Cultures. The London Meetings 2012-2014, edited by Brendan Larvor, 189–214. Cham: Springer International Publishing.

Hart, Herbert Lionel Adolphus, and Leslie Green. 2012. The Concept of Law, edited by Joseph Raz and Penelope A. Bulloch. 3rd edition. Oxford: Oxford University Press.

Kitchin, Rob, and Martin Dodge. 2011. Code/Space: Software and Everyday Life. Cambridge, MA: MIT Press.

Shapiro, Stewart. 2006. Vagueness in Context. Oxford: Oxford University Press.

Schauer, Frederick. 2013. "On the Open Texture of Law." Grazer Philosophische Studien 87(1): 197–215.

van Deemter, Kees. 2010. Not Exactly: In Praise of Vagueness. Oxford: Oxford University Press.

Waismann, Friedrich. 1945. "Verifiability." Proceedings of the Aristotelian Society, Supplementary Volume XIX: 119–150.