The ethics of belief in the context of data-driven knowledge

Emma Ruttkamp-Bloem[1]

[1] Department of Philosophy University of Pretoria; Centre for AI Research
emma.ruttkamp-bloem@up.ac.za

My general aim is to contribute to debates in data ethics around the trustworthiness of machine learning generated results. I analyse, in the context of critical machine learning, the conditions (norms) under which machine learning generated predictions or decisions generate epistemic beliefs. My analysis focuses specifically on engaging with debates in the context of the ethics of belief in order to firstly offer a philosophical framework for the call from critical machine learning for fair unbiased machine learning pratices, and secondly to argue in response to the call that fair unbiased machine learning practices are epistemic just practices.

The ethics of belief is an approach to the doxastic actions of agents situated at the intersection of epistemology, moral philosophy, philosophy of mind and psychology (Chignell 2018). The central question is whether belief acquisition, representation, communication (maintenance of belief), and revision (relinquishment of belief) are in some sense governed by norms. The father of the ethics of belief debate is the 19th century Cambridge mathematician and philosopher William Kingdon Clifford. Clifford's principle states that "It is wrong always, everywhere, and for anyone to believe anything on *insufficient evidence*". This principle is not only confined to describing the state (doxastic attitude) we are in when we form a belief, but also stretches to cover all our epistemic activities over time (Chignell 2018). We are obliged to go out to gather evidence and always have to remain open to new evidence (ibid.). The diachronic version of CP is: "It is wrong always, everywhere, and for anyone *to ignore evidence that is relevant to his beliefs, or to dismiss relevant evidence in a facile* way" (Van Inwagen 1996, 145).

As doing our "doxastic best" (ibid.) is both a moral and an epistemic issue (e.g. Locke 1690, Clifford 1877, Peirce 1877), my aim is to suggest here that one type of value governing belief formation in the context of data driven AI and fair unbiased ML is a value of epistemic justice. Epistemic justice as a value (at least partly) grounds doxastic norms because it speaks to an aspect of the foundation needed for generating just beliefs. It is not the only kind of value grounding doxastic norms but I argue that it is one of the most basic ones in the context of machine learning because it can give rise or exacerbate the harms from bias in machine learning. My intuition is that if the method (ML practices) giving rise to belief acquisition (epistemological commitment to the outcomes of machine learning practices) is not trustworthy, then there is some imbalance between the moral and epistemic values driving our belief acquisition. In the context of data driven AI, I want to illustrate one context within which such an imbalance can occur by linking epistemic unjust practices to the harm that can come from bias in machine learning.

The crux of Clifford's argument is the strong connection between the epistemic and the moral types of norm at play in his argument: The reasoning here seems to be as follows (ibid.): (P1) We have an epistemic obligation to possess sufficient evidence for all of our beliefs**; (**P2) We have a moral obligation to uphold our epistemic obligations; (C) Thus, we have a moral obligation to possess sufficient evidence for all of our beliefs. In terms of (P1) I argue that structural bias makes for insufficient evidence. But in addition, at a second level, insufficient evidence implies 'uncontextual' prediction, in the sense of not spelling out either the constraints within which predictions are generated by ML models, nor the constraints within which predictions should be interpreted or acted upon. (P2) gives the link between moral and epistemic values. I suggest a sub-argument for (P2) by considering the harms from decision-making systems in the context of structurally biased data. Then, by linking such harms to versions of epistemic injustice, I argue that in the context of critical machine learning the interplay between moral and epistemic norms is core to ensuring just machine learning practices, as the harms from machine learning imply epistemic unjust contexts of data gathering which may be at least one reason for insufficient (in the sense of biased) evidence in the first place. I conclude by affirming Clifford's conclusion when I show that *it is only in morally appropriate contexts that sufficient evidence can be generated*.

Clifford's argument in the context of critical machine learning then becomes: (P'1) Our epistemic obligations relate to ensuring sufficient evidence for beliefs. A person's knowledge is worthy of belief when there are *"reasonable grounds* for trusting" their veracity, knowledge and judgment (Clifford 1844, 46). Those grounds can only exist in a ML context if data in use has been gathered impartially, in a morally justifiable context.

(P'2) We have a moral obligation to uphold our epistemic obligations, as the latter can only be upheld if data is impartial; and data can only be impartial – and thus evidence be sufficient – if gathered in just circumstances. Epistemic just knowledge practices is at least a necessary condition for enabling doxastic agents to generate sufficient evidence for their beliefs. Fair classification practices (Crawford 2017) will guarantee representing cultural and historical divisions in society based on sensitivity to "relations of power and privilege that sustain injustice" (Mohanty1993, 53).

(C') Thus, we have a moral obligation to ensure knowledge is worthy of belief, i.e. that we believe on sufficient evidence, where 'sufficient evidence' refers both to fair data practices informing machine learning practices and to the clear articultation of the constraints, or ceteris paribus conditions, under which machine learning generated predictions or decisions are implemented.

I conclude that doxastic attitudes in the context of data-driven AI can only be generated via honouring the moral obligation to uphold (among others) the epistemic obligation to believe only on sufficient evidence (fair and unbiased data).

Keywords: Critical machine learning, Ethics of belief, Moral and epistemic norms, Bias, Accountability, Epistemic justice, Allocation and representational harm

References

1. Chignell, A.: "The Ethics of Belief", The Stanford Encyclopedia of Philosophy (Spring 2018 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/spr2018/entries/ethics-belief/ (2018).

2. Clifford, W.K.: “The Ethics of Belief”, in T. Madigan, (ed.), *The Ethics of Belief and other Essays*, Amherst, MA: Prometheus, 70–96 (1877 [1999]).

3. Clifford, W.K.: *The Scientific Basis of Morals and Other Essays* Gutenberg Ebook 2015 #50189 (1884).

4. Code, L.: Review of Miranda Fricker, Epistemic Injustice: Power and the Ethics of Knowing. *Notre Dame Philosophical Review*s, 3 (2008).

5. Crawford, K.: *The Trouble with Bi*as. NIPS Keynote (2017). https://www.facebook.com/nipsfoundation/videos/1553500344741199/ (last accessed 2019/10/15).

6. Fricker, M.: Epistemic Injustice: *Power and the Ethics of Knowing*. New York: Oxford University Press (2007).

7. James, W.: *The Will to Believe and other Essays in Popular Philosophy*, New York: Dover Publications, 1–31 (1896 [1956]).

8. Locke, J.: *An Essay concerning Human Understanding*, Oxford: Clarendon (1690 [1975]).

9. Mohanty, S.: The Epistemic Status of Cultural Identity: On “Beloved” and the Postcolonial Condition. *Cultural Critique*, 24:41-80 (1993).

10. Peirce, J.C.: The Fixation of Belief. *Popular Science Monthly* 12 (1):1--15 (1877 [1981]).