

Ignorance, insights and understanding in big data
practices:
An inferentialist account

María del Rosario Martínez Ordaz*
martinezordazm@gmail.com

Words: 459

The incorporation of big data into the empirical sciences has modified the ways in which scientific knowledge is pursued and achieved. Despite the positive outcomes of such incorporation, due to the increase in the amount of data there is a significant increase in the scientists' ignorance regarding the ways in which such data hangs together. In particular, much of the success in big data practices depends on computational processes that cannot be fully scrutinized, examined and justified by human agents (see Humphreys 2009). This lack of epistemic access to the ways in which the received data holds together when generating certain outputs is called *ignorance of theoretical structure* and it limits the understanding of the inference patterns that hold within a set (or a collection of sets) of data (Cf. Martínez-Ordaz 2020, forthcoming).

Furthermore, take scientific understanding to “consist of knowledge about relations of dependence. When one understands something, one can make all kinds of correct inferences about it” (Ylikoski 2013: 100). Now, when facing certain types of epistemic opacity, scientists would not be able to identify the relations of dependence between neither their beliefs nor the facts that they know; and this would have a significant impact on their achievement of understanding in big data contexts.

Here I submit that the ignorance that emerges when using big data in the empirical sciences is ignorance of theoretical structure *with reliable consequences* and I explain how this ignorance relates to the production of specific insights and later on, to further epistemic achievements such as knowledge and understanding.

I proceed in four steps.

- First, I address the connections between epistemic opacity in big data context and ignorance, and to do so, I adopt an inferentialist account of ignorance (Cf. Magnani 2009; Arfini 2016; Arfini, Bertolotti and Magnani 2018; Martínez-Ordaz 2020, forthcoming).

*Federal University of Rio de Janeiro, BRA.

- Second, I argue that some of the epistemic opacities that emerge in big data practices cause that scientists, irremediably, ignore the ways in which such outcomes were actually obtained. Here I am concerned with processes associated to the realization of algorithms in code as well as to the ways in which programs are actually run in particular instances (Cf. Humphreys 2009, Creel 2020).
- Third, I provide the generalities of an inferentialist account of understanding that could be explanatory of the partial overcoming of ignorance of theoretical structure in big data practices (Hazlett 2017). Here I also explain how, in approaches of this type, while *insights* do not suffice for understanding, they work as fix nodes within a structure, playing a crucial role in the constraining of the possibility space associated to the object of study and the later achievement of 'modal understanding' (Le Bihan 2017, Martinez-Ordaz forthcoming).
- Finally, I illustrate this with a case study from observational cosmology.

References

1. Arfini S (2016): "Thought experiments as model-based abductions". In: Magnani L, Casadio C (eds) *Model-based reasoning in science and technology: logical, epistemological, and cognitive issues*. Springer, Switzerland.
2. Arfini, S., T. Bertolotti and L. Magnani (2018): "The Antinomies of Serendipity How to Cognitively Frame Serendipity for Scientific Discoveries", *Topoi* 39: 939–948.
3. Baumeister AA, Hawkins MF, López-Muñoz F (2010): "Toward standardized usage of the word serendipity in the historiography of psychopharmacology" *J Hist Neurosci* 4(1):91–98.
4. Copeland S (2015) *The case of the triggered memory: serendipitous discovery and the ethics of clinical research*. PhD Thesis: Dalhousie University, Department of Philosophy.
5. Creel, K.A. (2020): "Transparency in Complex Computational Systems", *Philosophy of Science*, 87 (4).
6. Floridi, L. (2012): "Big data and their epistemological challenge", *Philosophy & Technology* 25(4): 435–437.
7. Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese*, 169(3), 615-626.
8. Leonelli, S. (2014): "What difference does quantity make? On the epistemology of Big data in biology", *Big data & Society* 1(1): 1-11.

9. Magnani L (2009): *Abductive cognition: the epistemological and ecocognitive dimensions of hypothetical reasoning*. Springer, Berlin.
10. Martínez-Ordaz, M. del R. (2020): “The ignorance behind inconsistency toleration”, S.I. Knowing the Unknown, *Synthese*.
11. Martínez-Ordaz, M. del R. (forthcoming): “Is there anything special about the ignorance involved in big data practices?”, in Lundgren, B. and N. Nuñez *TBD*, Philosophical Studies series, Springer.