

# From ignorance to trust in big data contexts

María del Rosario Martínez-Ordaz\*  
martinezordazm@gmail.com

For a variety of causes, scientific information is often inaccurate, poorly empirically supported, and not as relevant as it should be. And although there are good reasons for still aiming for accurate, empirically supported and relevant information in the sciences, the defective character of scientific data is not only ubiquitous but inevitable. However, while the presence of defective information in science tends to be naturally seen as part of the dynamics of scientific development, it is a fact that the larger the defectiveness of the information that scientists work with, the less justified they are in trusting such information.

In light of the above, there has always been a need for explaining under which circumstances scientists can, rationally, trust defective information in the sciences; yet, in recent years, this need has only increased due to the incorporation of novel epistemic practices into the scientific activity. In particular, during the last decades, the design of new technological and formal resources has allowed scientists to receive, order, and integrate enormously large amounts of data. Big data is the field that concerns the use of this kind of datasets -whose size is beyond the ability of typical database software tools to capture, analyze, store, and manage (Cf. Manyika et al., 2011). And while the incorporation of big data in different disciplines has come with a considerable amount of success and high levels of trust in the results of big data practices, it has also problematized the explanation and preservation of scientific rationality in defective contexts.

What I address here is the source of the scientists' trust in the products of big data, and whether such trust is justified. My main thesis can be summarized by the following two observations:

- Even if, at a certain moment, scientists cannot fully access and reconstruct the paths that are followed in the computer processes (particularly in scenarios in which deep learning techniques are involved), they can still trust *rationally* the outcomes of these processes. This is, if they ignore the warrant that is behind the computational procedures that are followed in their disciplines, they still are justified in trusting the resulting products.

---

\*Department of Logic, Nicolaus Copernicus University in Toruń / Federal University of Rio de Janeiro.

- The justification of this trust comes from the quality of the outcomes themselves; in particular, the information that the outputs contain is significantly *veracious* to the point in which it increases the empirical adequacy of the entailed measurements, descriptions, predictions, and explanations.

In order to explain this in more detail, I proceed in two major steps. First, I identify the type of ignorance that underlies epistemic practices in big data contexts, I contend that the type of ignorance that underlies big data practices is *ignorance of theoretical structure* (Cf. Martínez-Ordaz, forthcoming). Second, I explain how we make sense of the continued trust placed by scientists in defective information in the sciences consistently with ascribing rationality to them. Here, I scrutinize the notion of veracity of the data and I relate it to a more philosophically familiar concept: *empirical adequacy* (Cf. Bueno 1997). I explain that the connection between these two notions is that close, that if empirical adequacy suffices for justifying the scientists' trust in theories or products that they don't fully understand, veracity should do a similar work in big data contexts. I illustrate the above with a case study from observational cosmology.

## References

- Bueno, O. (1997): "Empirical adequacy: A partial structures approach", *Studies in History and Philosophy of Science Part A* 28 (4):585-610.
- Martínez-Ordaz (forthcoming): "Is there anything special about the ignorance involved in big data practices?" in Lundgren, Bjorn, L. and N. Nunez-Hernandez (Eds.) *Philosophy of Computing*, Philosophical Studies Series, Vol. 143.
- Manyika, J., Chui, M., Brown, B., *et al.* (2011): "Big data: The Next Frontier for Innovation, Competition, and Productivity". McKinsey Global Institute.