From serendipity and ignorance to knowledge and understanding in big data practices

María del Rosario Martínez-Ordaz
martinezordazm@gmail.com

Federal University of Rio de Janeiro, Brazil.

The incorporation of big data into the empirical sciences has modified the ways in which scientific knowledge was pursued and scientific methodology was followed. On the one hand, this incorporation has an impressively positive side providing scientists with extremely novel information. Nowadays, scientists are able to, for instance, reach objects that were initially inaccessible to them −either because of their size or because of their remoteness. On the other hand, the outcomes of big data applications often involve high degrees of epistemic opacity about how such outcomes were generated (Cf. Humphreys 2009, Floridi 2012, Leonelli 2014). This leaves scientists having to choose between rejecting the evidential role that these outputs may play or, at the risk of being irrational, relying on them even when ignoring where they come from and how they were obtained.

Here I aim at explaining how we can make sense of the continued trust placed by scientists in discoveries (made) by accidents while also ascribing rationality to them. I focus on the cases in which, in big data practices, scientists accidentally arrive at a game-changing novel result but they also, irredeemably, ignore the ways in which this result was actually constructed.

I argue that there is an extremely close relation between cases of serendipity in big data practices and a very particular type of ignorance: ignorance of theoretical structure *with reliable consequences* (Cf. Martinez-Ordaz 2020). I explain how this reliability of the outcomes makes possible the achievement of epistemic goods such as objectual knowledge of initially inaccessible objects as well as modal understanding of how these objects (could) behave and relate to one another, all this even if being ignorant of the inference patterns that govern the datasets from which the access to these objects is constructed.

I proceed in four steps.

- First, I address the connections between serendipity and ignorance, and to do so, I adopt an inferentialist account of ignorance (Cf. Magnani 2009; Arfini 2016; Arfini, Bertolotti and Magnani 2018; Martinez-Ordaz

2020). Here, I also address the importance of the game-changing character of serendipity and the possibility of interpreting and explaining, posthoc, the discovery (Cf. Copeland 2015, Baumeister *et al.* 2010).

- Second, I argue that some of the epistemic opacities that emerge in big data practices cause that, in some cases of discoveries (made) by accidents, 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 contend that the ignorance that underlies big data practices in the empirical sciences is ignorance of theoretical structure *with reliable consequences*; I explain the role that the game-changing character of serendipitous discoveries plays in assuring the reliability of these consequences.

- Finally, I explain how the reliability of serendipitous discoveries warrants the achievement of certain types of knowledge and understanding in big data practices. 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 episte-mology of Big data in biology", *Big data & Society* 1(1): 1-11.

9. Magnani L (2009): *Abductive cognition: the epistemological and ecocogni-tive dimensions of hypothetical reasoning.* Springer, Berlin.

10. Martínez-Ordaz, M. del R. (2020): "The ignorance behind inconsistency toleration", S.I. Knowing the Unknown, *Synthese.*