Is there anything special about the type of defectiveness involved in Big Data?
Some reflections on ignorance and Astrophysics[1]

Maria del Rosario Martinez Ordaz (UNAM)
martinezordazm@gmail.com

While philosophy of science has already started to discuss the different epistemic and philosophical difficulties that come with the use of Big Data in the scientific endeavor (Leonelli 2016), nothing has been said about the connections between the limited epistemic the epistemic role of ignorance when explaining scientific rationality in the era of Big Data. Here I aim at contributing to the philosophy of Big Data, in particular, to the epistemology of Big Data (and its applications in the empirical sciences). I focus on the epistemic role of ignorance when explaining the tolerance of defective information in the scientific practices of Big Data.

First, the use of defective (false, imprecise, inconsistent, conflicting, incomplete, etc.) information is ubiquitous in the empirical sciences. Second, it is well known, that the use of this type of information comes with the price of different degrees of ignorance (Wimsatt 2007, Norton 2008). Third, scientific rationality is only met either (a) when the degree of ignorance is lower than the degree of certainty –and thus, when sensible reasoning takes place– or (b) when the scientist does not endorse any doxastic commitments regarding the set of defective information. The combination of the above poses the following dilemma against scientific rationality: in cases of extreme degrees of defectiveness, either scientist are irrational for believing false information or they are irrational for reasoning under high degrees of ignorance –regardless their doxastic commitments.

The first horn of the dilemma is motivated thusly: It is commonly thought that anyone knowingly believing a falsity is irrational. It arguably follows that if a scientist believes some defective information (call this strong use of defective information), she is irrational. The second horn of the dilemma is this. Empirical scientists tend to use defective information only *pro tem*, and they expect uncertainty to decrease. Also, if the degree of defectiveness is higher than the degree of certainty, it is not clear which inferential procedures would allow for the preservation of the degree of certainty. So, anyone reasoning under high degrees of ignorance is irrational.

[1] To be presented at the *International Association for Computing and Philosophy – Annual Meeting 2019*. June 5-7, 2019, Mexico City, Mexico.

All this was, for many years taken for granted by philosophers and epistemologists of science, nonetheless, the current use of Big Data in some of our best scientific practices has presented a new challenge to our philosophical understanding of the role of defective information in the sciences. Nowadays scientific practice tends to use datasets whose size is beyond the ability of typical database software tools to capture, analyze, store, and manage (Manyika et al. 2011). Although much current scientific practice makes use of Big Data and scientists have struggled to explain precisely how Big Data do and machine learning algorithms actually work, they still rationally trust some significant chunks that these datasets contain. In order to explain the rationality behind Big Data-type of scientific practices, it is of prominent importance to provide a precise analysis of the nature of the defectiveness involved in this particular type of data sets.

Considering all the above, here I address an important question from the philosophy of Big Data, namely: *is there anything (epistemically) special about the type of defectiveness involved in Big Data?* I expect that the response to this question sheds light, in the long run, into the philosophical issue of the rationality behind the use of this type of defective information in the sciences.

In order to do so, I proceed in three steps:

- First, I claim that to pay special attention to the different degrees of ignorance that might surround the scientific use of defective information can be tremendously informative about the type of defectiveness of the information that scientists use and accept. In order to support this claim, I appeal to the explanatory success of similar analyses from the epistemology of science, according to which different degrees of defectiveness correspond to different degrees or types of ignorance that surround the relevant epistemic context (Wimsatt 2007, Norton 2008, Rescher 2009, Arfini 2019). Here, I focus on the standard understanding of the ignorance involved in the toleration of defectiveness in the sciences, namely, 'ignorance of truth value' (Unger 1975).

-Second, I present a case study that illustrates both the use of extremely large amounts of defective information as well as important degrees of ignorance. I focus on the particular case of Astrophysics as an exemplar of the use of defective information in the Big Data era. In Astrophysics, information of different types (such as images, redshifts, time series data, and simulation data, among others) is received in real-time in order to be captured, cleaned, transferred, stored and analyzed (Garofalo et al. 2016). The variety of the sources and the formats in which such information is received causes the problem of how to compute it efficiently as well as the problem of high dimensional data visualization, that is, how to integrate data that have hundreds of different relevant features. Since such datasets tend to increase in volume, velocity and variety (Garofalo et al. 2016), that makes it even harder to achieve any deep and exhaustive understanding of what they contain. However, this has not prevented astronomers from trusting important chunks of the information contained in such datasets.

-Third, I scrutinize the astrophysicists' use of defective information by appealing to the standard approaches to defectiveness via ignorance. I argue that in the case of Big Data in Astrophyiscs, while ignorance plays an important role for the understanding of defectiveness, 'ignorance of truth value' is not robust enough for being explanatory of this type of defects. Later on, appealing to the holistic properties of scientific bodies of knowledge, I characterize a distinct type of ignorance, namely, *ignorance of theoretical structure*, and I explain the epistemic practice in the case of Big Data and Astrophysics by apealing to it. Finally, I claim that the fact that ignorance of truth value is not enough for explaining the nature of defectiveness of Big Data tells something special about the type of defectiveness involved in Big Data.

(SELECTED) REFERENCES

Arfini, S. (2019): *Ignorant Cognition: A Philosophical Investigation of the Cognitive Features of Not-Knowing*, Studies in Applied Philosophy, Epistemology and Rational Ethics.

Ayyub B. M. ():"On uncertainty in information and ignorance in knowledge" *International Journal of General Systems*, Volume 39, 2010 - Issue 4, pp. 415-435.

Batterman, R. (Ed.) (2013): *Oxford Handbook for the Philosophy of Physics*. Oxford: Oxford University Press.

Garofalo, M., A. Botta and G. Ventre (2016):"Astrophysics and Big Data: Challenges, Methods, and Tools", *Astroinformatics (AstroInfo16) Proceedings IAU Symposium* No. 325, pp. 1-4.

Henningsson S. and W. J. Kettinger (2016) "Understanding Information Systems Integration Deficiencies in Mergers and Acquisitions: A Configurational Perspective", *Journal of Management Information Systems*, Volume 33, 2016, pp. 942-977.

Herausgeber: Lenhard, Johannes, Carrier, Martin (Eds.) (2017): "Mathematics as a Tool: Tracing New Roles of Mathematics in the Sciences", *Boston Studies in the Philosophy and History of Science*, Springer.

Kelly, T.: 2002, 'The Rationality of Belief and Some Other Propositional Attitudes'. *Philosophical Studies* 110 (2) 163-196.

Leonelli S (2016). *Data-Centric Biology: a Philosophical Study*. Chicago, IL, Chicago University Press.

Manyika, J., Chui, M., Brown, B., et al. (2011) *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute.

Norton, J. (2008):"Ignorance and Indifference", *Philosophy of Science* 75 (January) pp. 45–68. Copyright 2008 by the Philosophy of Science Association.

Olum, K. (2004): "Conflict between Anthropic Reasoning and Observation", *Analysis* Vol. 64, No. 1 (January), pp. 1-8.

Rescher, N. (2009): *Ignorance: (On the Wider Implications of Deficient Knowledge)*, University of Pittsburgh Press.

Rovane, C. (2004). "Rationality and Persons", In Piers Rawling & Alfred R. Mele (eds.), *The Oxford Handbook of Rationality*. Oxford: Oxford University Press. pp. 320--342.

Siegel, H.: 1985, 'What is the question concerning the rationality of science?' *Philosophy of Science*, pp. 517–537.

Šešelja D. and Straßer C. (2014) Epistemic Justification in the Context of Pursuit: A Coherentist Approach, *Synthese*, 191, pp. 3111-3141.

Unger, P. (1975): *Ignorance: A Case for Scepticism*, Oxford University Press.

Wimsatt ,W. C. (2007): *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*, Harvard University Press.