Is there anything special about the ignorance involved in big data practices?

María del Rosario Martínez-Ordaz

Federal University of Rio de Janeiro
Programa Nacional de Pós-Doutorado PNPD/CAPES

martinezordazm@gmail.com

June 18, 2020.
COLÓQUIO VIRTUAL SELF - PROBLEMAS FILOSÓFICOS

COLÓQUIO **SELF** *Problemas Filosóficos*

Here, I content that

The ignorance that emerges when using Big Data in the empirical sciences is quite unique: it is a limit case of *ignorance of theoretical structure with reliable consequences*.

Plan

1 Preliminaries

2 Why to care about ignorance in BD?

3 Ignorance about what?

4 Ignorance and BD

What is there
to be studied

preliminaries

why

ignorance

igno -BD

Preliminaries

Data

Anything that can be soundly recorded in a relational database respecting semantic and pragmatical requirements.

Big Data

The field that concerns ways to work with datasets whose size is beyond the ability of typical database software tools to capture, analyze, store, and manage (Cf. Manyika et al., 2011).

| volume, velocity, variety, veracity and value.

Note

The name Big Data does not only indicate the amount of data that is managed but, more importantly, the range of computational methods used to work with such data .

Preliminaries

Good news!

Different scientific disciplines that has been benefited from the emergence of Big Data and data science; this has happened for sciences such as astronomy, geology, climatology and biology which have a long history of working with large datasets.

Astronomy
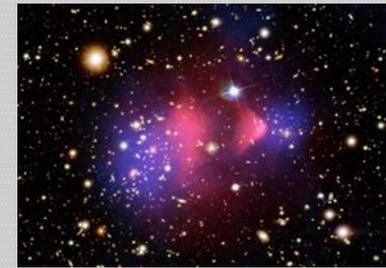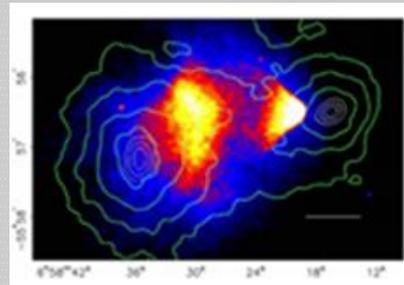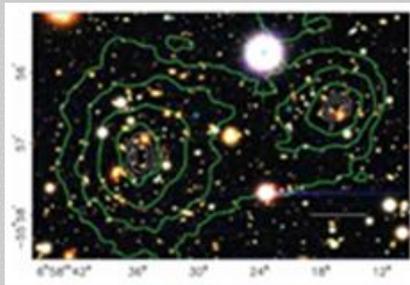
Allowed us to *see* something that was initially inaccessible

Why to care?

Is the amount a problem? I

Yes, there is an obvious exponential growth of data on an ever-larger number of topics, but complaining about such overabundance would be like complaining about a banquet that offers more than we can ever eat (...) We are becoming data-richer by the day; this cannot be the fundamental problem (Floridi 2012: 436).

Correlations

Is the amount a problem? II

As a matter of fact, the novelty of Big Data, at least for epistemology of science, should not lie in the sheer quantity of data involved, but rather in (1) the prominence and status acquired by data as commodity and recognized output, both within and outside of the scientific community and (2) the methods, infrastructures, technologies, skills and knowledge developed to handle data. (Leonelli 2014: 2)

What happens to the agents?

Why to care?

Some problems?

The incorporation of these new resources to any scientific discipline has come with some problems. In particular, with the increase in the amount of data comes an increase in the degree of **ignorance**.


What happens to the agents?

Ignorance about what

Ignorance

Absence of knowledge.

Facts/alethic value

1. Factual ignorance

Something, somewhere

2. Objectual ignorance
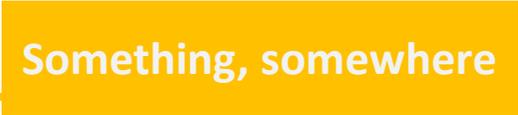
3. Procedural ignorance — **How to do something**

4. Ignorance of theoretical structure

Inference patterns

Ignorance about what

Ignorance of theoretical structure

Lacking knowledge of the (relevant) inference patterns that scientific theories allow for. When ignoring (the relevant parts of) the theoretical structure of a theory, scientists are not capable of grasping abstract causal connections between the propositions of their theory, they can neither identify the logical consequences of the propositions that they are working with nor can explain under which conditions the truth value of such propositions will be false. (Martínez-Ordaz 2020: 12)

Ignorance and BD

Some problems:

Telescopes   Observatories

Digitized data

Processing pipeline

Information/ knowledge

Storage   Storage   Storage

Storage   Dedicated archives   Storage
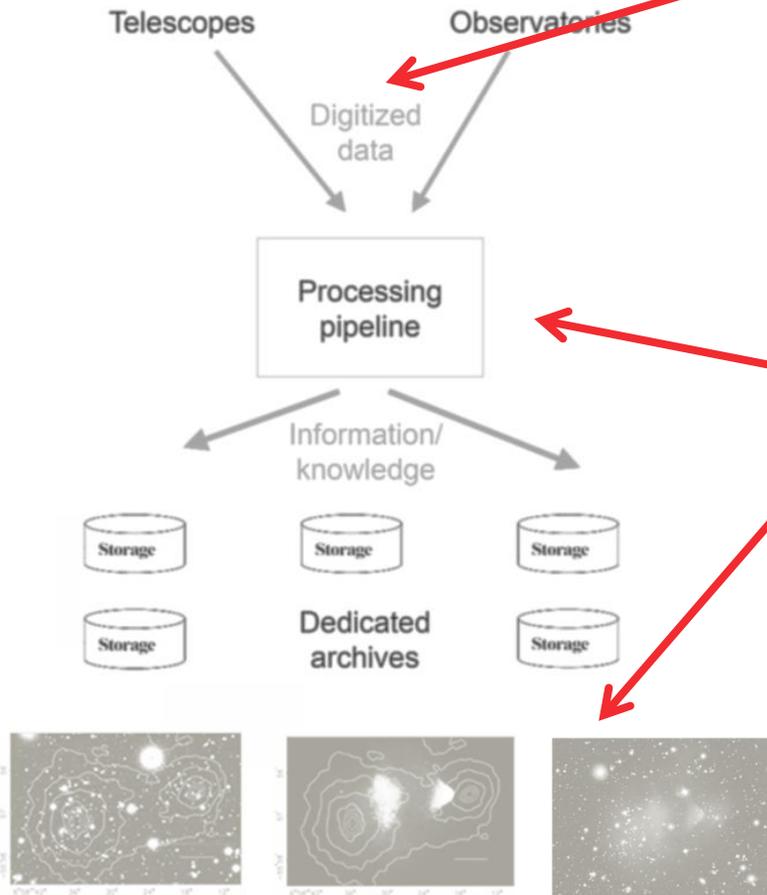
Not knowing the value of the data (Cf. Floridi 2012).

Correlations → ¬(causal exp) (Mayer-Sch• onberger and Cukier 2013: 14).

**Epistemic opacity
-- products
-- processes**

No understanding

Defective information

Ignorance and BD

Not knowing the value of the data (Cf. Floridi 2012).	**ignore** the specifics of the domains of application
Correlations → ¬(causal exp) (Mayer-Sch• onberger and Cukier 2013: 14).	**ignorance** of causal links
Epistemic opacity **-- products** **-- processes**	Factual **ignorance** Procedural **ignorance**
No understanding	**ignorance** of causal links
Defective information	Factual **ignorance**
Object of study	Objectual **ignorance**

Particularities

(a)

The human agents' cognitive limitations prevent them from computing, classifying and analyzing by themselves the data at the speed it is received.

(b)

The massively cooperative nature of Big Data practices makes the transmission and acquisition of knowledge very opaque processes.

Ignorance and BD

ignore the specifics of the domains of application

ignorance of causal links

Factual **ignorance**

Procedural **ignorance**

ignorance of causal links

Factual **ignorance**

Objectual **ignorance**

Instruments

Machine implementation

How all these fit together?

Theories

Ignorance of theoretical structure

preliminaries

why

ignorance

igno -BD

Ignorance and BD

ignore the specifics of the domains of application

ignorance of causal links

Factual **ignorance**

Procedural **ignorance**

ignorance of causal links

Factual **ignorance**

~~Objectual **ignorance**~~

Instruments

Machine implementation

How all these fit together?

Theories

Ignorance of theoretical structure

COLÓQUIO
SELF
Problemas
Filosóficos
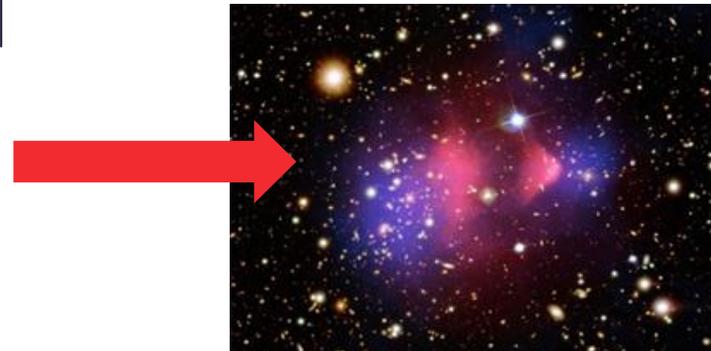
Final remarks

1. Still a lot to explore (agent-level)

2. Ignorance and BD: theoretical structure + reliable consecuences

3. Is our trust justified?

Obrigada!

Is there anything special about the ignorance involved in big data practices?

María del Rosario Martínez-Ordaz

Federal University of Rio de Janeiro
Programa Nacional de Pós-Doutorado PNPD/CAPES

martinezordazm@gmail.com

June 18, 2020.
COLÓQUIO VIRTUAL SELF - PROBLEMAS FILOSÓFICOS

preliminaries

why

ignorance

igno -BD