

## AGPhil 6: Symposium: Epistemology of Big Data in Physics IV

Zeit: Freitag 10:30–12:30

Raum: GW2 B2900

AGPhil 6.1 Fr 10:30 GW2 B2900

**Data science and explanatory power** — ●SERGEY TITOV — Institute of Philosophy, Russian Academy of Sciences, Moscow

The analysis-available data has grown immensely in the past decades, it has led us to the new type of research called "data-intensive". This research design mostly relies on vast amounts of data and use of complicated (commonly non-parametric) statistics. Such data analysis techniques show impressive results in predicting phenomena or its characteristics (for example, climate models) but suffer from serious loss in explanatory power (Calude & Longo, 2015). In some cases, models which are generated by nonparametric methods on given data are so complex, that it is nearly impossible to understand its structure. This problem is contemplated from philosophical and mathematical points of view. From philosophy's standpoint authors provide new structures of science which either take data-intensive research into account (Pietsch, 2015) or are fully based on it (Napoletani & Panza, 2011). Second approach attempts to explore mathematical foundations of this loss in explanation power. Calude and Longo in their work (Calude & Longo, 2015) use Ramsey theory and demonstrate that some of the patterns found in data-intensive research may be caused only by size of dataset and nothing else.

This work gives all-round view on this problem and tries to analyse some of the data-intensive researches in the manner described above.

AGPhil 6.2 Fr 11:00 GW2 B2900

**Causation, probability and all that: Data science as a novel inductive paradigm** — ●WOLFGANG PIETSCH — Munich Center for Technology in Society, TU München, Germany

Some have claimed that genuine data science is impossible since inductivism has allegedly long been refuted as a sound scientific methodol-

ogy. Instead, I argue that data science stands in an old and venerable empiricist tradition which includes highly influential scientists and methodologists like Francis Bacon, Isaac Newton or John Stuart Mill. An inductive methodology is sketched that could serve as a fundamental conceptual framework for data science. On this basis, I disentangle the conceptual muddle behind the claim that correlation replaces causation in data science, which is often held to be one of the central tenets in this discipline. By contrast, causation remains the crucial concept to distinguish between meaningful and accidental correlations, i.e. between those that allow for prediction and manipulation and those that do not.

AGPhil 6.3 Fr 11:45 GW2 B2900

**Do the Beads Still Need a String? Old and New Challenges for Turning (Big) Data into Evidence** — ●JOHANNES LENHARD — University of Bielefeld

How are data turned into evidence? Finding answers to this question is a central endeavour of epistemology and methodology of the sciences. One classic 19th century locus is the dispute between John Stuart Mill and William Whewell about the nature of scientific knowledge. Does it grow with the amount of data available, or is it of a general character that is based on something different than data? Whewell remarked that collecting observational data is not sufficient, since "the beads still need a string."

Recent claims about data-driven research (DDR) seem to challenge this viewpoint. DDR allegedly heralds a new paradigm in which the data themselves can replace also the string.

My talk critically assesses recent disputes and claims about DDR and big data. One strategy is by historical comparison to earlier, pre-computer examples.