Studies in History and Philosophy of Science 59 (2016) 1–10

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

journal homepage: www.elsevier.com/locate/shpsa

Graph-based inductive reasoning

Marcel Boumans

Utrecht University School of Economics, Kriekenpitplein 21-22, 3584 EC Utrecht, Netherlands

ARTICLE INFO

Article history:
Received 3 February 2016
Received in revised form
18 April 2016

Keywords:
Expert judgment
Graphical method
Inductive reasoning
Measurement
Statistical method
Visual cognition

ABSTRACT

This article discusses methods of inductive inferences that are methods of visualizations designed in such a way that the "eye" can be employed as a reliable tool for judgment. The term "eye" is used as a stand-in for visual cognition and perceptual processing. In this paper "meaningfulness" has a particular meaning, namely accuracy, which is closeness to truth. Accuracy consists of precision and unbiasedness. Precision is dealt with by statistical methods, but for unbiasedness one needs expert judgment. The common view at the beginning of the twentieth century was to make the most efficient use of this kind of judgment by representing the data in shapes and forms in such a way that the "eye" can function as a reliable judge to reduce bias. The need for judgment of the "eye" is even more necessary when the background conditions of the observations are heterogeneous. Statistical procedures require a certain minimal level of homogeneity, but the "eye" does not. The "eye" is an adequate tool for assessing topological similarities when, due to heterogeneity of the data, metric assessment is not possible. In fact, graphical assessments precedes measurement, or to put it more forcefully, the graphic method is a necessary prerequisite for measurement.

1. Introduction

The twentieth century witnessed an exponential growth of social statistics. At its beginning, the magnitude of data was in the order of a few kilobytes; half way through that century it was in the order of megabytes, and at the end of the twentieth century it was in the order of gigabytes, thereby earning the label of Big Data (see Diebold, 2013); in other words, each 50 years the magnitude of social statistics increased by a factor 1000.[i] Despite this enormous expansion of data, the nature of the core problem of inductive inference remained the same: How to infer meaningful patterns from these masses of data when there is little or no theory to guide the inferences and there are not yet any standardized objective procedures to follow? Part of the answer is that one needs additional expert judgments.[ii] But then the subsequent question is—when and in what manner are these expert judgments instrumental?

This article will discuss methods of inductive inference that are methods of visualizations designed in such a way that the "eye" can be employed as a reliable tool for judgment. I use the term "eye" as a stand-in for visual cognition and perceptual processing. The reason is that I will discuss literature from the beginning of the twentieth century in which the term "eye" was commonly used instead of the currently used concept of visual-cognitive system. The method of visualizations that will be discussed is the method of graphs, also called the method of curves, that was developed around 1900.[iii]

In her position piece, Annamaria Carusi (2012) advocates that more attention should be paid in philosophy of science to the epistemological role of visualizations: "[I]t is necessary to understand how vision works embedded in *epistemic* contexts, as playing

E-mail address: m.j.boumans@uu.nl.

[i] Although these numbers are not so impressive as they are in natural science, where they are in the order of petabytes, the additional "big" problem is that social data are much more heterogeneous.

[ii] See Boumans 2015 for an extensive substantiation of this argument.

[iii] Visual displays have of course various different roles in science, for example to make theories comprehensible, to present data sets, or to analyze data. This article focuses on the latter.

http://dx.doi.org/10.1016/j.shpsa.2016.05.001

a crucial role in the formation of evidence for claims" (p. 107).[iv] Notwithstanding the bulk of studies in cognitive sciences which has accumulated in the past few decades, attention to the visual as such is still scarce in philosophy of science today. According to Carusi, these studies challenge philosophy of science to re-think its position on three key distinctions: the qualitative/quantitative, the subjective/objective, and the causal/non-causal distinction. This article hopes to contribute to re-thinking the first two distinctions.

A good overview of studies in cognitive science on the use of visualization in scientific thinking and reasoning is provided by Hegarty (2011).[v] Maria Hegarty summarizes some of the main advantages that visual displays afford for cognitive tasks: 1. External storage of information; 2. Organization of information; 3. Offloading of cognition on perception, and 4. Offloading cognition on action. This article will focus on the second and third tasks of visual displays. It will show that in the organization of information—here called visualization—the "eye" also plays a crucial role. The organization of information requires not only "vision to think," that is, the offloading of cognition on perception, but also "vision to visualize."

Halfway through the twentieth century, visualizations came to imply the involvement of computers, and were often called simulations.[vi] Although in that period visualizations were only possible with analog computers,[vii] in economics the digital computer was considered as an instrument of observation, supplying "a viewing equipment to the economist in a manner analogous to the microscope for biologists" (Shubik, 1960). According to Mary Morgan (2004, p. 363), who discusses Martin Shubik's account of simulation, Shubik may have inherited this metaphor of the microscope from Oskar Morgenstern (1954). Morgenstern made this comparison with the microscope for cases in which one has to deal with an enormous amount of data for which there is no theory that could attach any meaning to data. Without theory, one would be "just looking" or "merely looking." In an earlier stage of a science this may lead to data of a new kind. When the telescope and microscope were invented, "all that mattered was to take these wonderful new instruments and to look, to look practically anywhere. Some phenomena would turn up. Totally unsuspected, be they the moons of Jupiter or some tiny amoebae in a drop of water" (1954, p. 540).

While Morgenstern spoke of "just looking" when making observations with a microscope, Shubik noted rightly that, as every user of a microscope would admit, the metaphor of a microscope implies a "specimen" to be observed with the microscope, and that the setting up of such a specimen would require "a great amount of work" (Shubik, 1960, p. 908). In social science, however, these specimens are not material objects but

> models of those things. Thus, where the biologist prepares slides so that they can "see" certain things with the microscope, economists prepare models so that the relevant parts of the world they specify can be "read" by the computer. And if the models prepared for the economists' computer-as-microscope are not natural, they are, of course, artificial constructions, man-made. (Morgan, 2004, p. 364)

Both Shubik and Morgan observe that inductive inference from an enormous data set is not "just looking" at the "raw data" but it requires a specific kind of preparation—a visualization—to have something to look at. Morgan calls this kind of preparation—rightfully in my view—modeling, in the sense of making a representation.[viii]

The analogy between specimen and model also implies an analogy between the epistemic problems on both sides. As much as a specimen can lead to artefacts caused by the method of preparation, a model built to make a phenomenon visible may also create artefacts about it. The problem however is that the epistemic strategies to distinguish between valid facts and artefacts, such as control of possible confounding effects and systematic error, replicability, data reduction and calibration do not travel very well from the material world of specimens to the virtual world of models. Although the problem of artefacts is a relevant issue, it will not be further discussed in this paper because it will go beyond the topic of this paper and is already extensively discussed elsewhere, see for example Boumans, 2002 and Franklin, 1997.

To explore the preparation of "specimens" which enables the observation of social phenomena, it is useful to move to the period when these kinds of modeling strategies were discussed most explicitly. Firstly, I will discuss the general shared epistemology of the method of graphs around 1900. Second, I will discuss a specific and rich case in which the method of graphs has been used to infer meaningful patterns from the enormous amount of available data. This case study concerns how Warren M. Persons designed a "barometer" to inform businessmen and politicians about the "general business conditions." The last section will discuss—as an illustration—a current case of pattern recognition that applies a similar strategy based on the epistemological advantages of the "eye" to organize information.

2. The method of graphs

Histories of the graphic method (Funkhouser, 1937; Hankins, 1999; Klein, 1995; Maas and Morgan, 2002) point to William Playfair as the originator of this method for social data. In the introduction of his (1796) *A Real Statement of the Finances and Resources of Great Britain*, Playfair mentions explicitly the advantage of using graphs and charts, which he called "lineal arithmetic":

> [I]t is to give a more simple and permanent idea of the gradual progress and comparative amounts, at different periods, by presenting to the eye a figure, the proportions of which correspond with the amount of the sums intended to be expressed. As the eye is the best judge of proportion, being more accurate and quicker than any other of our organs, it follows, that where-ever *relative quantities*, a gradual increase or decrease of any revenue, receipt or expenditure of money, or other value, are to be stated, this mode of representing it is peculiarly applicable, as it gives a simple, accurate, and permanent idea; it produces form and shape to a number of separate ideas, which are otherwise abstract and unconnected; for in a numerical table there are as many distinct ideas given, and to be remembered, as there are sums. The order and progression, therefore, of those sums, are also to be recollected by another effort of memory, while this unites proportion, progression, and amount, all under one simple impression of vision, and consequently one act of memory. (Playfair, 1796, pp. v-vi)

Although a century earlier René Descartes, one of the originators of analytic geometry, had stated that "imagination or visualization

[iv] See Mößner, 2015 for a similar position.

[v] I thank one of the anonymous referees for pointing me at the relevant literature.

[vi] See Morgan, 2004 for an insightful discussion of simulations as a new "technology."

[vii] See Boumans 2012 for a treatment of analog computers as tools of intelligibility.

[viii] This is more explicitly stated and explored in Section 5.iii of Morgan 2012, which carries the title "Specimens = Models."