Studies in History and Philosophy of Biological and Biomedical Sciences 66 (2017) 63–72

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

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

Galton, reversion and the quincunx: The rise of statistical explanation

André Ariew[a,*], Yasha Rohwer[b], Collin Rice[c]

[a] Department of Philosophy, University of Missouri, Columbia, MO 65211, United States
[b] The Department of Humanities and Social Sciences, Oregon Institute of Technology, Klamath Falls, OR 97601, United States
[c] Department of Philosophy, Bryn Mawr College, Bryn Mawr, PA 19010, United States

ARTICLE INFO

Article history:
Received 27 February 2017
Received in revised form
14 August 2017
Available online 27 October 2017

Keywords:
Francis Galton
Quincunx
Statistics
Explanation
Normal distribution
Causation

ABSTRACT

Over the last six decades there has been a consistent trend in the philosophy literature to emphasize the role of causes in scientific explanation. The emphasis on causes even pervades discussions of non-causal explanations. For example, the concern of a recent paper by Marc Lange (2013b) is whether purported cases of statistical explanation are "really statistical" or really causal. Likewise, Michael Strevens (2011) argues that the main task of statistical idealizations is to distinguish between the causal factors that make a difference to the phenomenon to be explained and those that do not. But, the philosophy literature poorly reflects the history of the development of statistical explanation in the sciences. Francis Galton's (19th century) explanation for the laws of heredity is our case. Galton's statistical explanation was both innovative for his time and influential to our contemporary sciences. The key points to understanding Galton's statistical explanation for reversion is that it is autonomous from the real-world biological properties that make up an instance of reversion while still being approximately true of many real-world biological phenomena. Ours is an expanded discussion of ideas originated in Hacking (1990) and Sober (1980). We will articulate these features and compare our account with that of Lange and Strevens.

Published by Elsevier Ltd.

> "The typical [statistical] laws are those which most nearly express what takes place in nature generally; they may never be exactly correct in any one case, but at the same time they will always be approximately true and always serviceable for explanation" (Galton, 1877, p. 17)

1. Current trends in the philosophy of scientific explanation are causal centric

As Michael Strevens (2012) reminds us, a great preponderance of philosophical work in the area of scientific explanation has been focused on the study of causal relations. The motivation for causal approaches can be traced back to issues with Carl Hempel and Paul Oppenheim's account (1948) according to which many scientific explanations involve deductions from premises stating a generalization and some initial conditions. But, as the well-known flagpole case demonstrates, Hempel and Oppenheim's deductive account fails to adequately account for the explanatory role causal facts play. We can derive the height of the flagpole from the length of its shadow in the way required by Hempel and Oppenheim's account, but the unpalatable result is that the cause being explained by the effect. Therefore, there is a problem calling derivations of flagpole heights from shadows genuine cases of explanation. In contrast, deducing the shadow from the height of the flagpole is a genuine explanation because it cites causes to explain their effects.

According to Strevens, ever since the flagpole case, much of the literature in philosophy of scientific explanation has set its agenda accordingly to the following three questions:

1. What are causal relations?
2. How do deductive derivations and other semantic apparatus in science represent causal relations?
3. Besides the need to represent the causal production of the explanandum, what other norms govern the construction of scientific explanations?

This agenda has produced an intellectually unsatisfactory literature in the philosophy of scientific explanation in several regards. Here are three.

First, Strevens' agenda neglects important scientific advances that ought to guide philosophers' discussions about the norms of

* Corresponding author.
E-mail addresses: ariewa@missouri.edu (A. Ariew), Yasha.Rohwer@oit.edu (Y. Rohwer), crice3@brynmawr.edu (C. Rice).

https://doi.org/10.1016/j.shpsc.2017.08.001
1369-8486/Published by Elsevier Ltd.

scientific explanation. These important advances are typically neglected because they do not easily conform to popular philosophical accounts of good scientific explanation. Our case study is just one instance: Francis Galton's statistical explanation for what he called the "law of heredity": the processes of heredity maintain a fixed distribution of variation across generations. We would expect the explanation for a biological phenomenon to reference biological mechanisms and processes. But, Galton's explanation for this phenomenon is surprising because his explanation only makes reference to a mathematical result. The intergenerational stability of the distribution of variation is a deductive consequence of the distribution of variation of the previous generation. It matters not whether the ensemble is composed of coin tosses, shots on target, heights of soldiers, or biological characters, if the frequency of characters in the ensemble is normally distributed (under ordinary or equilibrium conditions) then it can be deduced that in the next generation there will be a normal distribution of the same mean and dispersion. Consequently, the exceptional characters will revert as a deductive consequence of the normal distribution (Hacking, 1990, p. 183).

Galton's explanation was a first in biology. Galton's novel use of statistics to explain a real-world phenomenon led him to develop the statistical techniques of correlation, linear regression, and a variety of standards for goodness of fit between data and theory, all still used by scientists. According to historian Stephen Stigler, Galton's work "represents the most important step in perhaps the single major breakthrough in statistics in the last half of the nineteenth century" (Stigler, 1990, p. 281). Further, Galton's mathematical theory of inheritance was the basis for the 20th century synthesis of Darwinian natural selection and genetical theories of inheritance. It also laid the groundwork for the development of modern population and quantitative genetics (Fisher, 1953, p. 5). For such an important scientific advance, one would think that philosophers of science would pay more attention to it. But, while the overall number of citations referring to Galton's techniques exceeds that of the flagpole literature by orders of magnitude, the opposite is true within the philosophy of scientific explanation literature.[1]

A second problem with the tradition that stems from Hempel and the flagpole is that it has brought an over-emphasis on the role of causes in scientific explanation.[2] The current trend in the literature on non-causal explanation is concerned with a kind of demarcation exercise: to determine whether purported cases of explanation are genuinely non-causal or only appear to be non-causal but are in fact truly causal. Examples include a mother dividing 23 strawberries among three children (Lange, 2013a); why it is impossible to walk through Königsberg and cross each of the seven bridges once and only once (Pincock, 2007); how to explain the periodicity of cidada life cycles (Baker, 2005); and the nature of the differences between genetic drift and natural selection in the modern theory of natural selection. (Lange, 2013b), (Walsh, Lewens, & Ariew, 2002). We believe the demarcation question is largely irrelevant to understanding the nature of Galton's statistical explanation for reversion and we shall criticize Marc Lange's (2013b) work accordingly.

The key points to understanding Galton's explanation for the law of heredity is that the statistical explanation is autonomous from the real-world biological properties that make up an instance of reversion while still being approximately true of many real-world biological phenomena. Galton's statement in the epigraph is a good expression of the view. Galton's reversion explanation is autonomous in the sense that the mathematical features of a statistical equation that Galton cites are sufficient to explain biological reversion, even though the statistical equation fails to accurately represent any real-world events that make up any particular instance of reversion.[3] The idealized statistical explanation of reversion is sufficient on the condition (or to the degree) that the frequency distribution of the trait in the population is *approximately* normal. What determines the degree to which a real world population approximates a normal distribution depends on *minimal material* requirements of the system, namely that the ensemble is the result of numerous randomized trials whereby the probability of the outcome of the one is independent of the probability of the outcome of any other. We discuss the details below. To help clarify our position we will contrast our account of the nature of Galton's reversion explanation with that of Marc Lange's account of what makes regression explanations "really statistical" as opposed to latently causal.

A third problem with the entrenched Hempelian tradition is that there is too much emphasis on univocal and general accounts of scientific inquiry. We ought to recognize the possibility that there exist multiple adequate accounts of good scientific explanation. We will criticize Michael Strevens (2016) account of the role of idealizations in scientific explanations in this regard. His exclusive focus on causal-difference makers fails to account for the autonomy of statistical explanation. The point of Galton's use of the deductive features of a statistical law to explain regression is not, as Strevens has it, to highlight causal-difference makers. Rather, the point is that the statistical laws do all the explaining without the need to refer to any causal features of the ensemble.

We won't extrapolate from the historical case to formulate a universal account of all statistical explanations.[4] Rather we hope to learn some valuable lessons about best-case practices—in this case, Galton's explanation for reversion—in order to extract some norms of scientific explanation. We mean to reset the project of the philosophy of scientific explanation by freeing it from the tradition that automatically thinks about flagpoles and causal relations. Our historical case refers not to philosophers, like Hempel and Oppenheim, thinking generally about science, but a scientist inventing a unique way of explaining a natural phenomenon.[5] To us, Galton's explanation is exemplary: it represents a norm of scientific explanation whose nature is not revealed by answering the three questions Strevens lays out for us. Ultimately we want to understand how the deductive properties of an idealized representation of a biological phenomenon can be, in Galton's words, "approximately true and always serviceable for explanation" of a real-world biological process. To fully appreciate Galton's innovation it is important to note the state of statistical methodology at the time of Galton's writing. As you shall see, by articulating Galton's response to the pioneers of the statistical methodology we will have a deeper understanding of the aspects of Galton's explanation we are highlighting.

[1] Exceptions include: Hacking (1990), Sober (1980), Lipton (2009), Lange (2013a, 2013b), Gayon (1998), Ariew, Rice and Rohwer (2015), Radick (2011).

[2] Even Strevens' third question is dependent on an account of causation because in order to articulate the "other norms" we need to know what distinguishes them from the norm of representing causal production.

[3] We adopt Ian Hacking's (1990) concept of "autonomy" account for the nature of the Galton's innovative statistical explanation. See also Sober (1980) and Ariew et al. (2015).

[4] We consider other cases in science in Rice, Rohwer, Ariew (in prep.), "Explanatory Schema and the Process of Model Building".

[5] We certainly don't claim originality for this way of doing philosophy of science. The works of Nancy Cartwright are exemplary.