Quantitative realizations of philosophy of science: William Whewell and statistical methods

Kent Johnson

Department of Logic and Philosophy of Science, UC Irvine, 3151 SSPA, Irvine, CA 92697, USA

ARTICLE INFO

ABSTRACT

In this paper, I examine William Whewell's (1794–1866) 'Discoverer's Induction', and argue that it supplies a strikingly accurate characterization of the logic behind many statistical methods, exploratory data analysis (EDA) in particular. Such methods are additionally well-suited as a point of evaluation of Whewell's philosophy since the central techniques of EDA were not invented until after Whewell's death, and so couldn't have influenced his views. The fact that the quantitative details of some very general methods designed to suggest hypotheses would so closely resemble Whewell's views of how theories are formed is, I suggest, a strongly positive comment on his views.

© 2011 Elsevier Ltd. All rights reserved.

> When citing this paper, please use the full journal title *Studies in History and Philosophy of Science*

1. Introduction

A distinctive feature of the empirical sciences is that their methods are typically quantitative. As a given discipline matures and develops, this quantification of methods tends to increase. At the same time, many methodological theories within the philosophy of science are presented largely or entirely verbally. It is commonplace, for instance, to describe what the scientist does at a high level of generality, so that the low-level quantitative issues are abstracted away from. Given the very broad aims of such projects, this abstraction is often appropriate. Still, when all is said and done, expansive philosophical theories about the methodology of science should jibe with the quantitative methods that actually drive scientific research, at least to the extent that the former imply quantitative details. The purpose of this paper is to consider a very general view of how science works, due to William Whewell, in the light of the quantitative details of some common statistical methods.

This paper is organized as follows. §1 presents a brief overview of Whewell's views. In §2, we consider how these views relate to contemporary statistical methods. As I argue below, statistics is an excellent vantage point from which to evaluate Whewell's views. Unsurprisingly, the fit between Whewell's views and statistics is not perfect. Nonetheless, I shall argue, the correspondence is remarkably good. Moreover, we will see that Whewell's work provides an important corrective to certain contemporary scientific attitudes.

Before beginning, a couple caveats are in order. First, there is a certain ahistorical aspect to this project. I do not suggest that Whewell had statistics in mind when he was writing—importantly, I argue, quite the opposite. Nor do I suggest that statisticians and users of statistical methods learned to ply their trade by reading Whewell. Rather my aim is to offer a partial evaluation of the accuracy of Whewell's characterization of the workings of science. Second, I do not claim that Whewell's philosophy is the unique best representation of contemporary science. Some of what I will discuss admits of non-Whewellian interpretations; I claim only that Whewell's work describes it quite well. Although I do not argue for it here, I do not believe the same claim can be made for the views of many others, e.g. John Stuart Mill (Whewell's rival) or W.v.O. Quine.

2. Decomposition, colligation, explication: an overview of Whewell's views of theory formation

For much of the nineteenth century, William Whewell (1794–1866) engaged in an extraordinarily prolific academic career at Trinity College, Cambridge. He conducted scientific research in

mineralogy, the study of the tides, and political economy; he also wrote textbooks on mechanics, mathematics, and astronomy and physics (e.g., Whewell, 1825, 1836, 1856, 1819, 1838, 1833). Additionally, he wrote a large three-volume *History of the Inductive Sciences*, in which he investigated the development of various sciences (astronomy, optics, mechanics, electricity, zoology, physiology, etc.) from their ancient origin to their then current status (Whewell, 1858). Along with these many and varied interests, Whewell also developed and defended a sweeping view of the methodology of science in his two-volume *Philosophy of the Inductive Sciences* (Whewell, 1847. Unannotated citations will be to this work; e.g., II, 26 refers to page 26, of volume II of Whewell, 1847.).

Although Whewell's philosophy of science is complex, its core concerns his views about induction, which itself is centered around four main processes. Using his neologisms, they are: (i) the decomposition of facts, (ii) the explication of conceptions, (iii) the colligation of facts, and (iv) the verification of the resulting proposition (which includes his well-known consilience of inductions).[1] The first three processes concern the formation of a theory, and the last involves the confirmation of the theory. This paper concerns the first three processes; I reserve discussion of his views on the theory confirmation for another time.

The most central theme in Whewell's philosophy of science is his insistence upon the importance of the contribution of the scientist's mind in the process whereby 'Science is built up by the combination of Facts' (II, p. 26). Whewell held that there were a small number of fundamental ideas—e.g., space, time, number, force, motion, position, etc.—in terms of which the raw data of science invariably must be understood. For Whewell, ideas are 'general relations among our sensations, apprehended by an act of the mind, not by the senses simply' (II, p. 25). (Whewell's ideas should not be confused with Kant's pure concepts; below, we will consider several differences between them.) These ideas are so fundamental to science that particular subsets of them were taken to be nearly definitive of various scientific fields (e.g., II, pp. 116–117). To transform our brute perceptual sensations into perceptions of objects or other external phenomena, we must use our ideas to organize the sensations. For example,

[w]hen any one has seen an oak-tree blown down by a strong gust of wind, he does not think of the occurrence any otherwise than as a *Fact* of which he is assured by his senses. Yet by what sense does he perceive the Force which he thus supposes the wind to exert? . . .. It is clear upon reflection that in such a case, his own mind supplies the conception of extraneous impulse and pressure, by which he thus interprets the motions observed (II, p. 28; cf. p. 25).

Very often, the fundamental ideas will be too general and amorphous to be of direct use in actual scientific practice. What we must then do, Whewell argues, is refine and fine-tune these ideas so that they fit the particular scientific endeavor at hand. Whewell refers to the resulting refinements as 'conceptions':

[B]y the word *Idea* (or Fundamental Idea,) used in a peculiar sense, I mean certain wide and general fields of intelligible relation, such as Space, Number, Cause, Likeness; while by *Conception* I denote more special modifications of these ideas, as a *circle*, a *square number*, a *uniform force*, a *like form* of flower. (II, p. 380)

As we will see, finding the right conception is a centerpiece of Whewell's philosophy of science (e.g., II, pp. 5–26).

The psychological component of science is so important that a 'certain activity of the mind is involved, not only in seeing objects

erroneously, but in seeing them at all' (II, p. 29). Regarding the related question of whether we can separate our ideas (or conceptions) from the external facts, Whewell is firm: we cannot (e.g., I, p. 36; II, pp. 26–33; 47). The idea of force is too fundamental to seeing the oak tree blown down by the wind; the idea of space is too fundamental to seeing a physical object (II, p. 29); the ideas of resemblance and difference are too fundamental to classificatory endeavors such as botany (II, p. 367); and so on. Indeed, attempting to view facts without the interpretive guidance of ideas 'leaves the mind overwhelmed, bewildered, and stupefied by particular sensations, with no means of connecting the past with the future, the absent with the present, the example with the rule; open to the impression of all appearances, but capable of appropriating none' (II, p. 47).

If there were nothing more to say about the relationship between the facts in the world and the ideas in the mind, there would be little reason to have much faith in the results of science. We may have organized the data using the wrong ideas, the wrong refinements of ideas, non-intellectual notions based on fear, admiration, etc. Fortunately, this is not the case.

We are not able, nor need we endeavor, to exclude Ideas from our Facts; but we may be able to discern, with perfect distinctness, the Ideas which we include. We cannot observe any phenomena without applying to them such Ideas as Space and Number, Cause and Resemblance, and usually, several others; but we may avoid applying these Ideas in a wavering or obscure manner, and confounding Ideas with one another (II, p. 31).

Although ideas and facts are inextricably intertwined, we can still make a study of which ideas we have used, and how we have used them. The ongoing process of refining and studying our conceptions is for Whewell the major component of scientific progress.

For Whewell, conceptions (and fundamental ideas—following Whewell, I will frequently not distinguish the two) play two important roles. First, they provide us with 'the most universal, exact, and simple' conceptions which we use to 'decompose' the raw facts we encounter into the more tractable data upon which science is built (cf. the second and third Rules; II pp. 32–33). This process is the *decomposition of facts*:

Thus the Facts which we assume as the basis of Science are to be freed from all the mists which imagination and passion throw round them; and to be separated into those elementary Facts which exhibit simple and evident relations of Time, or Space, or Cause, or some other Ideas equally clear. We resolve the complex appearances which nature offers to us, and the mixed and manifold modes of looking at these appearances which rise into our thoughts, into limited, definite, and clearly-understood portions. (II, p. 33)

Once the facts have been decomposed with the aid of the appropriate conceptions, we can then record the actual measurements. E.g., after we decide to study the stars with reference to their number, relative positions and distances, rather than, say their participation in various astrological configurations, we can engage in the practical task of determining the former magnitudes (e.g., II pp. 337–338). For our purposes, we may take this aspect of his philosophy of science as relatively straightforward.

Thus far, we've seen the importance for Whewell of carefully analyzing which ideas we are using, and how we are using them—i.e., what particular conceptions we have refined these ideas into. In fact, Whewell argues that discovering and understanding the fundamental ideas and their subsequent conceptions is 'the most important step' in induction (II, p. 383; cf. II, pp. 51, 91). Whe-

---

[1] Whewell himself lists six processes that enter into the 'formation of science' (Whewell, 1847, II, p. 336), but the four listed above receive by far the most emphasis.

well thinks that this 'psychological' component of science—the *explication of conceptions*—is often underestimated. He complains that the ideas and conceptions of science often seem so obvious that they appear paltry next to the deductive work that utilizes them: 'men often admire the deductive part of the proposition, the geometrical or algebraical demonstration, far more than that part in which the philosophical merit really resides' (II, p. 91). Similarly, he rails against those who disparage earlier failed attempts to find a precise, useful idea/conception: 'It is as if a child, when its teacher had with many trials and much trouble prepared a telescope so that the vision through it was distinct, should wonder at his stupidity in pushing the tube of the eye-glass out and in so often' (II, p. 378; cf. pp. 60 ff.). More generally, Whewell frequently comments on the great difficulty of discovering a suitable idea/conception: 'The process of obtaining new conceptions is, to most minds, far more unwelcome than any labour in employing old ideas. The effort is indeed painful and oppressive; it is feeling in the dark for an object which we cannot find' (II, p. 101; cf. also II, pp. 7, 8, 15, 46, 55–57, 376–379).

The second role that properly explicated conceptions plays is far more difficult and crucial to Whewell's philosophy of science. As he himself immediately notes, it 'by no means follows that when we have thus decomposed Facts into Elementary Truths of observation, we shall soon be able to combine these, so as to obtain Truths of a higher and more speculative kind' (II, p. 34). Instead, after we have decomposed the facts, we must find the right conceptions to bind them back together. In Whewell's terminology, we must *colligate the facts*:

> Facts are bound together by the aid of suitable Conceptions. This part of the formation of our knowledge I have called the *Colligation of Facts*: and we may apply this term to every case in which, by an act of the intellect, we establish a precise connexion among the phenomena which are presented to our senses (II, p. 36).

A colligation 'binds together' the various diverse facts (I, p. 43; II, pp. 27, 36, 50, 60), creating a 'bond of unity' (II, 35, 46). Colligating the facts with an appropriate conception is like stringing a collection of pearls together to form a necklace (II, pp. 48, 52). Thus, colligation involves a 'step of a higher order' (II, 34). As Whewell often emphasizes (II pp. 11–16, 379), the explication of conceptions and the colligation of facts are intimately related. The former 'must be carried on with a perpetual reference to' the latter (II, p. 379; cf. p. 12). This makes sense: In order to find a properly explicated conception, we need to keep an eye on how we might pull the basic facts together into the kind of unity that can be had only by a true scientific theory. But in order to pull the facts together in this way, we must keep an eye out for the kind of conception(s) that can do the job. In other words, a conception that does not bear on the facts is empirically vacuous, and a collection of facts that are not organized in any way is too overwhelming, misleading, and complicated to be of any real use. Moreover, this 'feedback loop' between explication and colligation constitutes much of how science progresses. A reasonably good explication of a conception can be the basis of a colligation of facts, which in turn can point the way to an even more precise explication, which leads to a better colligation, and so on. (And of course, we may even be led to decompose the facts in some more accurate way.)

Not only are explication and colligation tightly connected, they are 'the two processes by which we arrive at science' (II, p. 5), and collectively 'they constitute the mental process of *Induction*; which is usually and justly spoken of as the genuine source of all our *real general knowledge* respecting the external world' (II p. 46). Unlike many other philosophers, induction for Whewell is not a mere summary and generalization of the facts. Instead, as he frequently

stresses, by using 'superinducing' conceptions upon the facts together, induction always imports something further into the data: 'the particular facts are not merely brought together, but there is a New Element added to the combination by the very act of thought by which they are combined' (II, p. 48; cf. pp. 53, 77, 85, 88, I 25). When this 'new element' that is added to the combination is part of a true theory, the propriety of colligating the facts with the given conception(s) becomes retrospectively obvious. Indeed, as time goes by, it is hard to imagine the particular facts in any fashion than the one supplied by the conception(s) used (II pp. 8, 48, 52).

Conceptions, we have just seen, play a crucial role in both the decomposition and subsequent colligation of facts. Thus, it is puzzling that Whewell would associate their explication so much more strongly with colligation than with decomposition (e.g., II, pp. 12, 46, 379, 50, 53, 54, 379, 383). Why is the difficulty and importance of properly explicating the conceptions used in the colligation of facts emphasized so much more than the explications used to decompose them? A reasonable answer might be that, although the decomposition of facts into elementary truths is of fundamental importance to science, it is generally not nearly as difficult or frequent as the cycling between explications and colligations. Although it may be difficult for the botanist to arrive at a suitable definition (or even a conception) of a rose (II, pp. 424–425), she nonetheless can make various sorts of relative and absolute measurements of various candidate plants. E.g., she can measure number of petals, length of stamen, etc., building of course upon a prior theory of seeds, stamen and the like (which themselves depend on prior decompositions, explications, and colligations). All this is not to say that the decomposition of facts is intrinsically easy, but that Whewell may have arranged things so that the hard part could be located in the colligation of previously decomposed facts.

The part of Whewell's philosophy of science that we have just reviewed concerns the 'discovery' or formation of a theory. As we've seen, theory formation for Whewell is a rational, inferential process. It is, however, quite distinct from the confirmation of theories: 'The Invention of the Conception was the great step in the *discovery*; the Verification of the Proposition was the great step in the *proof* of the discovery' (II, p. 51). In this sense, it is quite distinct from the various forms of hypothetico-deductivism which begin by generating definitions and axioms, and deriving empirical consequences from them, which can be checked against the world. For Whewell, determining the exact conceptions to be used, and coming to understand their exact nature constitutes an enormous amount of the scientific enterprise. Indeed, Whewell held that the right conceptions are determined towards the *end* of the discovery of the theory, not at the beginning. Moreover, the conception may not be statable as a precise definition. Similarly, the proper colligation of facts is at the end of the discovery, and it too may not be statable as a formal axiom. For further discussion of Whewell's views of theory formation, cf. e.g., Buchdahl (1991), Fisch (1985, 1991), Ruse (1991), Snyder (1997a,b), Snyder (2006, 2008), Yeo (1993). Those familiar with this literature will notice that my reading of Whewell is considerably closer to the view developed by Snyder than to that of others, e.g., Fisch. While I stand by my interpretation (e.g., I do not consider it plausible to read Whewell in a non-realist, conventionalist fashion), rational minds can nevertheless disagree.

Finally, despite the crucial and irremovable contribution of the mind, Whewell's scientific realism is firm. The scientist 'may *understand* the natural world, but he cannot *invent* it' (II, p. 379; cf. II, pp. 7–8). Similarly, 'Man is the Interpreter of Nature, and Science is the right Interpretation' (I, p. 37). Indeed, Whewell's realism is one half of his 'fundamental antithesis of philosophy' between our ideas and the external world: '[w]ithout Thoughts there could be no connex-

ion; without Things, there could be no reality' (I, pp. 17–18). Whewell's realism is grounded in his theology. The fundamental ideas are God's ideas, and they reflect how He chose to structure the universe. In His infinite beneficence, He gave humans these ideas too, so that they could understand and appreciate His creation.

(The reader may have noticed some similarities between Kant's transcendental philosophy and Whewell's philosophy of science. While Whewell freely admitted Kant's influence on his thinking, these similarities should not be overstated. For our purposes, we can observe three crucial differences between Whewell's fundamental ideas and Kant's pure concepts. First, unlike Kant, Whewell thought that there were more Fundamental Ideas yet to be discovered. Second, Whewell's Fundamental Ideas correctly represent objective features of the external world (cf. Snyder, 2006, pp. 42–47).). Third, Whewell's ideas lack any Kantian apriricity regarding their application. Although they are necessary for experience and knowledge about the world, they can be incorrectly applied, resulting in a flawed theory, or misperception. E.g., 'A vague and loose mode of looking at facts very easily observable, left men for a long time under the belief that a body, ten times as heavy as another, falls ten times as fast (II, pp. 37–38). Indeed, we've seen that a central aspect of Whewell's philosophy involves the difficult and ongoing process of understanding, via properly explicated conceptions, just how the ideas relate to the world.)

While there is much more to Whewell's philosophy of science, we have seen his views on how theories are 'discovered'. An obvious question is how accurately he characterized this process, which is the topic of the next section.

3. Whewell's views as realized in contemporary statistics

3.1. Justification of using statistics to evaluate Whewell

In this section, I consider how well Whewell's general picture of science is realized in the quantitative details of contemporary statistical methods. Statistics is a good field to evaluate Whewell's views from. Among academic disciplines, statistics plays a unique role as both a freestanding academic discipline and also a clearinghouse for a great deal of the methodologies used in the other sciences, including the hard sciences like physics and chemistry. Indeed, statistics encompasses an enormous amount of scientific methodology quite generally, and to a far greater extent than any other field. (E.g., Volume I of the widely respected 'Kendall and Stuart's' *Advanced Theory of Statistics* series begins by defining: 'Statistics is the branch of scientific method that deals with the data obtained by counting or measuring the properties of populations of natural phenomena' (Stuart & Ord, 1994, p. 2).) Thus, a general theory of how science works, such as Whewell's, should make close contact with statistics.

There is a second, historical, reason for examining Whewell's views from this standpoint, which concerns that fact that he had little to say about statistics. Whewell formed his views by examining an enormous amount of other scientific work (cf. Snyder, 2008, pp. 217–221 for an interesting discussion of this last point). Since statistics did not directly influence his philosophy in the way that, e.g., Newton's physics did, the former can act as a largely independent source of data against which Whewell's claims may be checked. As Whewell would put it, his philosophy makes novel predictions about a field different in kind from those he used to form the theory. Thus, predictive accuracy in this case approaches the strongest form of confirmation of the theory (beyond, of course, a consilience from multiple such different fields) (II pp. 62–65).

Initially, it might seem overstated to say that statistics had little influence on Whewell's thinking. He does discuss some roughly statistical ideas such as the 'method of means', the 'method of least squares' and the like (II pp. 395–412). Moreover, by Whewell's time, Jakob Bernoulli, DeMoivre, Laplace, Gauss and others had established certain basic elements of probability theory (e.g., Stigler, 1986, 1990). In fact, Whewell also helped to form the Statistical Section of the British Association for the Advancement of Science, as well as the Statistical Society of London (Snyder, 2008, p. 166). Perhaps, then, statistical methods had more of an impact on Whewell's thinking than I just suggested?

A bit of thought removes this worry, for three reasons. In the first place, the statistical methods of Whewell's time, such as they were, had not made their way from Continental Europe. As the historian of statistics Anders Hald notes:

> When [Galton] began his statistical work in the 1860s, the methods of Laplace and Gauss and their followers were not generally known in Britain. Galton therefore developed his own crude methods, numerical and graphical, for analyzing normally distributed observations in one and two dimensions. Although his methods were primitive, his ideas were clearly expressed and had a profound effect on the development of the British Biometric School. (Hald, 2007, p. 135).

Since the last edition of Whewell's *Philosophy of the Inductive Sciences* was in 1860 (Whewell died in 1866), it wouldn't have been possible for him to take advantage of the 'crude' methods available in Britain. Similarly, the prominent statistician Bradley Efron has noted that '[t]he current era is the first century in which statistics has been widely used for scientific reporting' (Efron, 1986, p. 1). Given the particular approach we will adopt below, this latter comment is perhaps even more relevant.

Secondly, the last century has seen a massive development of the field of statistics. In terms of the sheer quantity of statistical methods, as well as their relative mathematical and computational sophistication and intensity, contemporary statistics bears little resemblance to the 'statistics' of Whewell's day. Lacking the enormous contributions of R. A. Fisher, Jerzy Neyman, Karl Pearson, Egon Pearson, Charles Spearman, L. J. Savage, and many others, statistics for Whewell was largely devoted to recording large tables of data, sometimes also including some of the most rudimentary descriptive statistics of the sample, such as its average on each measurement. To take just two examples, as late as the first half of the twentieth century, there were several huge shifts in the scientific community regarding what should be counted as a *sample*, perhaps the most fundamental notion of statistical theory. (Desrosières, 1998, pp. 210–211). Similarly, the mathematical foundations of probability, given in the Kolmogorov Axioms, appeared over 65 years after Whewell's death. Whewell's time lacked such basic elements as coefficients of reliability or correlation, confidence intervals, goodness-of-fit tests, etc.

Finally, the techniques discussed below are all part of an area known as Exploratory Data Analysis (EDA). EDA is designed not for the purposes of the statistical testing of hypotheses or making statistical inferences, but for suggesting hypotheses which later on might be subject to various sorts of confirmatory tests. EDA is a fairly recent area of research in statistics, partly because these computationally intensive techniques are difficult to perform without a relatively powerful computer. All of the techniques discussed below were not part of statistical methodology until well after Whewell's death.[2] In short, Whewell's statistics bears about as much resemblance to the contemporary field—and to EDA in particular—as

---

[2] Principal component analysis was introduced to statistics by Pearson (1901); the singular value decomposition of a complex rectangular matrix was proven to exist and to have its statistically most desirable properties by Eckart & Young (1939) (cited in Horn & Johnson, 1985, p. 426; cf. Stewart, 1993 for earlier proofs of the decomposition of certain special subclasses of matrices). Finally, independent component analysis was largely given its initial development by Comon (1994).

Newton's mechanics does to contemporary quantum mechanics. To the extent that these methods differ from anything Whewell was familiar with, and yet his characterization of how science works nonetheless *still* applies, Whewell's views are confirmed. One might also evaluate Whewell's views in an opposing direction, and focus on methodological details that have been incorporated into our current, vastly more complex, techniques. Here one might credit Whewell to the extent that he saw that they reflect an important and central aspect of science, as opposed to a more 'local' technique that really only applied to the methodological tools of the day. For example, minimizing the ordinary least squared deviation between a model and the data is still a common optimization criterion (cf. Forster, 1988); it is, however, by no means the only one (e.g., for non-Gaussian distributions, it differs from the maximum likelihood estimate, and the latter is often to be preferred). A more critical evaluation, of course, would come from those aspects that Whewell incorrectly deemed unimportant or central to scientific activity.

The dual role of statistics—independent academic discipline, and repository of scientific methods—means that there are two ways we might use it. On the one hand, we might use it just as another scientific field. In such a case, we should look into the minds and findings of great statisticians, checking whether they engaged in the decompositions, explications and colligations that Whewell claims they do. On the other hand, if we take statistics, broadly construed, to encompass (a large portion of) scientific methodology, we might take a more general approach. Rather than checking whether Whewell's views are realized in statisticians' research—i.e., *developers* of statistical methods—we could examine whether they are found in the practices of ordinary scientists—i.e., *users* of statistical methods. Because of the very broad applicability of statistics in the sciences, if it turns out that it is standard statistical practice to engage in the processes Whewell described, we will have gone very far towards confirming Whewell's views about the nature of science.

3.2. Uncertainty as the fundamental idea of statistics

If we hope to find Whewellian thinking in statistical methods, we need to consider what fundamental ideas should be associated with the latter. Whewell was, as we've seen, adamant that each science have its own (not necessarily proprietary) ideas, which are then explicated into various more useful conceptions. But the ideas of space, time, and cause seem more appropriate to those empirical disciplines that use statistics, and less so to statistics itself. The idea of number is perhaps a somewhat better candidate, although it too, is often rather far removed from the actual study of statistics. E.g., in the theoretical populations often studied by statisticians, the number of elements within any given subpopulation is commonly either zero or infinite. A better choice of a fundamental idea of statistics, I suggest, is *uncertainty*[3]. At its heart, statistics involves the identification, presentation, analysis, management, and control of various types of uncertainty present in data and theories. Dealing with uncertainty is the central theme of statistics, whether it takes the form of determining the distribution(s) from whence the data came, the right kind of statistical test to perform and concomitant inference to draw, identifying the number of unobserved factors, components, dimensions, etc. underlying a data set and their relations to one another and the observed variables, or any of the other tasks commonly assigned to statistics. Moreover, although Whewell never considered uncertainty as a fundamental idea, it is still Whewellian in spirit. As he often noted, as various fields come into being, there will be new ideas and conceptions distinct from those already in use (II, pp. 18, 33, 39, 43, 88, 100).

Taking uncertainty as the primary fundamental idea of statistics also points the way towards the kinds of conceptions that it is refined into. The various ways of dealing with uncertainty (e.g., determining distributions, drawing statistical inferences, confidence intervals etc.) may be seen as conceptions drawn from the fundamental idea. We will see detailed examples of this below.

3.3. Statistical decomposition of facts

Let us now consider how well Whewell's views of theory formation are represented in statistics. In contemporary terms, the decomposition of facts is found in the theory of measurement, which can be usefully viewed from (the somewhat artificially distinguished) practical and theoretical perspectives.

From the practical perspective, the decomposition of facts is straightforward. When a scientist sets out to study some particular phenomenon, she must decide what kinds of data to gather. E.g., a geologist who is studying the geological composition of a region might begin by collecting samples from various locations in the region. But once she is back in her lab with a few hundred rock samples, the real data collection begins. The actual samples are enormously complex, and contain a great many features, only some of which are relevant to what the scientist is studying. Should she measure the amount of magnesium in calcite and/or the amount of sodium in muscovite? How about the sulfide content of the samples, the crystal size of the carbonates, the spacing of the cleavage, the elongation of the ooliths, tightness of the folds, and the number of veins and fractures per square meter in the sample? (This example is borrowed from Basilevsky, 1994, pp. 255–257.) Which of these features, along with many, many others, is relevant depends on the nature of the scientist's particular investigation. That is, the 'beginning of exact knowledge' involves the scientist's determination of the relevant properties of the samples to measure (II, pp. 33–34). This decomposition of the facts 'resolves the complex appearances' in the rock samples 'which nature offers to us', and exchanges the 'mixed and manifold modes of looking at these appearances which rise into our thoughts' for 'limited, definite, and clearly-understood portions' which can then be represented quantitatively (Ibid.) Indeed, the contribution of the scientist's conceptions to the data is so important that virtually every introductory statistics textbook emphasizes the importance of carefully planning the experiment before collecting any empirical samples. Preplanning is important because which conceptions will be needed to decompose the empirical samples crucially depends on the precise details of what the researcher is attempting to explore, and which conceptions are used can greatly affect which samples are obtained, how many are needed, and how they are obtained. (E.g., if fractures are relevant, those measurements may need to be obtained directly at the sites, before the rock samples are extracted from their natural location.) All this is just to say that statistical methodology requires a great deal of input from the scientist in the extraction of the raw scientific data (often a matrix of quantitative data, of n observations each measured along k many dimensions) from the bare empirical facts (e.g., a collection of rocks).

Whewell is also correct to stress the importance of carefully analyzing the conceptions employed at this stage. E.g., if the geologist also wants to know the hardness of the rock samples, she will have to determine which of several radically distinct conceptions of hardness are relevant to her study (cf. Wilson, 2006, pp. 335–355 for an interesting discussion of various different scientific measures of 'hardness').

---

[3] I use the rather general term *uncertainty*, rather than, say, *randomness*, so as to avoid any metaphysical commitments regarding the nature of stochastic phenomena. Talking of uncertainty also makes it easy to include such phenomena as measurement error, and the statistical study of deterministic systems under conditions of incomplete knowledge.

From the theoretical perspective, one also sees that the decomposition of facts requires ideas when we consider the scales of measurement associated with different kinds of data. To see this, consider the five most common scales.

An *absolute* scale is used when the number given by the measurement cannot be transformed into any other number. An example of this is count data; e.g., if the number of eggs a chicken laid in a month is 25, this number cannot be changed. Thus, an absolute scale is unique, and all numerical properties are preserved.

A *rational* scale is used when all measurements agree on what counts as zero units, but where the measurements can differ as to the size of the units. E.g., it is clear that something has zero length in feet if it also does in meters, but something's length in feet is 3.28 times its length in meters. A new rational scale S_2 can be created from a rational scale S_1 by the transformation $S_2 = bS_1$ ($b > 0$). Switching rational scales preserves ratios of pairs of measurements: $bx/by = x/y$.

An *interval scale* is used when measurements have distinct zero points, but agree on the ratios of given intervals between pairs of points (and order). E.g., the Fahrenheit and Celsius scales locate 0 degrees at different temperatures, but agree on how large the range is of the high and low temperatures on Tuesday compared to Wednesday. A new interval scale S_2 can be created from an interval scale S_1 by the affine transformation $S_2 = a + bS_1$ ($b > 0$). Switching rational scales preserves interval ratios:

$$\frac{(a + bx) - (a + by)}{(a + bz) - (a + bw)} = \frac{x - y}{z - w}.$$

An *ordinal* scale is used when measurements agree only on the ordering of the data. E.g., a grade of an A in a class is better than a B, but it cannot be inferred that the difference in quality of performance is the same as that between a B and a C. A new ordinal scale S_2 can be created from an ordinal scale S_1 by the transformation $S_2 = f(S_1)$, where f is any monotonically strictly increasing function—i.e., if a < b, then $f(a) < f(b)$. Switching ordinal scales preserves order: a < b iff $f(a) < f(b)$.

A *categorical* scale is used when the measurements agree only on whether the data fall into the same or different categories. E.g., coding males with 1 and females with 2 only indicates that the two categories differ. A new categorical scale S_2 can be created from a categorical scale S_1 by the transformation $S_2 = f(S_1)$, where f is any injection—i.e., if $a \neq b$, then $f(a) \neq f(b)$. Thus, switching categorical scales preserves only the identities of the categories.

The fact that there are different scales shows that we must decide what the numbers in our data set mean. Just as we do not see the force that blows the oak tree, but must impute it to the scene we witness, so too, we must impute the nature of the scale onto our numerical measurements. E.g., we do not 'see' that our data are (merely) ordinally scaled. Instead, treating them as such is an inference from our theoretical understanding of the relations between the magnitudes we have measured. We import the idea that our quantitative measurements contain certain kinds of information and not others, and 'superinduce' it upon the facts. We see this particularly clearly when we theorize overtly about the nature of the scale imposed upon the data. For example, if hardness is taken to be (operationally) defined by location on the Mohs hardness scale, then this scale is absolute. However, if hardness is identified with the measurement from a sclerometer, then the Mohs scale is merely ordinal in nature.

Finally, in line with my interpretation of Whewell, decomposing the facts—whether it takes the form of deciding what types of things to measure or what types of scales the measurements fall on—is, in actual scientific practice, often easier than the colligation of facts. Although conceptions are used in statistical decompositions, as we'll see below, this is typically nowhere near as demanding a process as the subsequent colligation. (All this is not to say that there is not a rich mathematical theory behind measurement (e.g., Krantz, Duncan Luce, Suppes, & Tversky, 1971), only normally the measurement issues just discussed are less difficult.)

3.4. Statistical explication of conceptions and colligation of facts

We now turn to the most central aspects of the 'discovery' component of Whewell's philosophy of science: the explication of conceptions and the colligation of facts. Do statistical methods 'superinduce' a conception, a 'new element' upon the decomposed facts in a 'step of a higher order', which 'binds the facts together' like 'pearls on a string', thus creating a 'bond of unity'? In fact, this is an elegant description of what occurs throughout statistics. In general, successful statistical models work by reorganizing the data so as to reveal important aspects of the true, unobserved structure of the data and their source. (In contrast to Mill's prohibition against unobservables, Whewell's acceptance of the latter is crucial: even in the simplest 'location' model that treats each datum as merely the mean deviated by some 'error', i.e., $x_i = \mu + \varepsilon_i$, the model posits an unobserved bipartite structure of x_i.) This is especially clear for those statistical methods that are routinely used to formulate and suggest—or 'discover', in Whewell's words—new hypotheses, which is the heart of EDA. As an example of this, let's examine one common such technique, principal components analysis (PCA).[4] (The following two paragraphs lean heavily on Johnson, 2007.)

The nature of PCA can be brought out with a simple example. Suppose we are examining the concentrations of three chemicals X, Y, and Z in a given region. One hundred groundwater samples are taken from the region, and the amounts of each of X, Y, and Z are recorded. When the data are plotted as points on three axes, they are distributed as in Fig. 1a below. Rather than being randomly dispersed, the data appear to be structured around a two-dimensional plane. This structure is in one sense a real surprise, as it is extraordinarily improbable that a random sample of unrelated measurements would ever yield such a pattern. (The boxes are scaled to a 1-1-1 ratio to visually present the correlations, as opposed to the covariances, of the three variables.)

It's the essence of the sciences not to ignore such patterns. A natural first step is to try to understand 'how much' of a pattern is there, and what its nature is. Obviously, the relative concentrations of X, Y, and Z appear related. From the geometric perspective of the cube, the fit of the data on the angled plane (cf. Fig. 1b) is fairly close. The planar surface lies at a skewed angle, so all three axes of the cube are involved. But if we used a *different* set of axes, we could view the data as organized primarily along just two axes. That is, suppose we replaced axes X, Y, and Z with three new axes, A, B, and C. (If we keep A, B, and C perpendicular to one another, we can think of ourselves as holding the data fixed in space, but rotating the cube.) Moreover, suppose that we choose the axes so that A is that single axis on which we find as much of the variation in the data as possible. If we wanted to represent as much of the variation in the data as possible with just one axis, A would be our best choice. It wouldn't perfectly reproduce all the information about X, Y, and Z, but it would capture a lot of it. Now suppose we fix the second axis B so that it captures as much of the remaining variation in the data as possible, after we factor out the variation that A captures. Together, A and B would determine a plane lurking in

---

[4] PCA is actually a very good example, because some researchers consider it inferior to factor analysis, in that only the latter, they claim, produces a 'model' with parameters to be estimated. However, as the discussion below shows, this attitude is too narrow, at least from the present philosophical perspective.
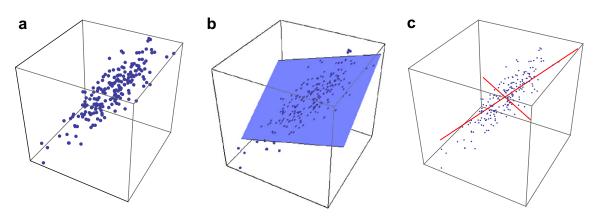
Fig. 1. (a) Some data measured in three distinct ways, corresponding to the three original axes. (b) The two-dimensional plane on which the data largely lie. (c) A new pair of (orthogonal) axes, which are linear combinations of the original ones, and which capture as much of the variance as any two axes can.

the three-dimensional space. (The two lines in Fig. 1c correspond to Axes A and B.) By projecting all the data onto this plane, we could recover much, although not all, of the information in the data. (We'll miss exactly that information regarding how far away from the plane the actual data points lie.) If we set axis C to best capture the remaining information, we will then be able to recover all the information in the original space. If, however, we decide to use only one or two axes, we can represent the data reasonably well in a less complex, lower dimensional, fashion.

The real scientific import of PCA comes when we find that just a few PCs can account for much of the variation in a lot of various measurements. For example, a number of studies of color and color perception have used PCA and related techniques to estimate the number of basis functions needed to replicate various data sets to a high degree of accuracy. E.g., Romney and Indow (2003) measured the amount of light reflected by each of 1269 color chips at 231 evenly spaced points in the visible spectrum. In short, they located these 1269 chips in a 231 dimensional space. They then estimated these chips' reflectance profiles using only the three 'best' dimensions. Despite this omission of 228 dimensions (i.e., 98.7% of them), the correlation between the estimates and the actual values was a striking .988. In other words, they obtained a resolution of 98.8% when $1269 \times 231 = 293{,}139$ data points were represented using only $(1269 \times 3) + (3 \times 231) = 4500$ numbers. Such patterns are far too extreme to be random, and they cry out for explanation. PCA and related techniques can help expose and quantify such patterns in useful ways. (Actually, this particular result of Romney and Indow's was obtained via some techniques closely related to PCA. However, I checked their data set using PCA, and got similar results. Using the best 3, 4, or 5 principal components captures 98.7%, 99.5%, and 99.8% of the (standardized) variation respectively.)

When the data are independently drawn from a multinormal distribution, it is even possible to conduct statistical tests to determine which PCs are statistically significant (e.g., Basilevsky, 1994, chapter 4). In short, a successful PCA can organize and re-present the data in such a way that allows us to derive an *explanandum*. Why should the 231 measurements of each of 1269 color chips behave (almost) as though they came from only 3 to 5 sources, instead of 231? At this point, a metaphysical/empirical hypothesis suggests itself: maybe they behave this way because there are only a few influences responsible for the reflectance profiles of the color chips. Typically, further research is performed to confirm or undermine such hypotheses.

The relationship between statistical techniques like PCA and Whewell's philosophy of science is straightforward. Quite simply, the former techniques, when used correctly, are quantitatively realized colligations of facts. By re-expressing the data in terms of PCs, we can, when successful, come closer to discovering the true struc-

ture underlying the original observations. For example, in the study of color mentioned above, the 'bond of unity' that 'binds together' the 1269 observations is the fact that nearly all of their variation in 231 dimensions can be captured with the smallest handful of carefully selected dimensions. More specifically, the bond of unity can be thought of as the three or so dimensions that capture virtually all of the statistical 'behavior' of the 1269 observations. Moreover, these newly discovered dimensions present us with 'Truths of a higher and more speculative kind' than we could ever hope to glean from the 293,139 original data points. (In fact, by performing an independently motivated rotation of the three axes, Romney and Indow discovered that they correspond to the physical properties of brightness, hue, and saturation.) Although Mill would not approve of the use of these PCs, because these latent variables are by their very nature unobservables, Whewell was right to suggest that theory formation, often with good reason, will posit them.

The statistical colligation just described makes crucial use of a conception derived from the fundamental idea of uncertainty. The conception employed is the much more specific criterion of a vector's best least squares fit of the data. That is, each successive PC removes as much of the remaining uncertainty, in the sense of unaccounted variance, as possible (subject to the orthogonality condition mentioned above). In a very straightforward sense, the 'particular facts' of the data set 'are not merely brought together, but there is a New Element', in the form of the new basis vectors 'added to the combination by the very act of thought' (i.e., statistical methods) 'by which they are combined' (II, p. 48; cf. also pp.77, 85, I 25). In a successful PCA, only a few vectors are retained (e.g., 5 of 231); thus, this 'New Element' can correspondingly also be thought of as a *removal* of something—irrelevant extra dimensionality—from the facts. By taking away these irrelevant dimensions, 'new truths are brought into view' (II, p. 43).

As Whewell predicts, determining the right conception to superinduce upon the facts is a difficult and ongoing process. In our PCA example, the first two PCs determine a plane that captures most of the information in the data. However, any two independent vectors on that plane can determine that plane. This means that the same amount of information can be recovered by any two independent linear combinations of the original manifest variables that lie on that plane. (Analogous remarks apply to any *k*-dimensional subspace of an *n*-dimensional vector space.) Thus, if we aim to find the 'true' axes that represent the two statistical factors that actually produce the behavior in the manifest variables we typically cannot accept the two extracted from the initial PCA. Although our conceptions from the PCA allowed us to locate the variation and uncertainty in the data in a plane, they must be further refined if we wish to discover a theory about the correct axes for the data.

The issue described above is one of a 'rotation' of the initial axes from the PCA to theoretically more satisfactory positions. What counts as 'theoretically more satisfactory' here depends a great deal on the researcher's background assumptions about the organization of the data. Sometimes a detailed background theory may dictate where the axes should be located, and the researcher will place them there 'by hand' with no further statistical guidance. Other times, considerably weaker background assumptions may motivate a statistically guided rotation. One may wish, for example, to reduce the complexity of the PCs by a 'varimax' rotation, and seek that position that most nearly approximates the case where each PC loads at nearly 1 or 0 for each variable—i.e. to each manifest variable, each PC contributes fully or not at all. Alternatively, one may seek to reduce the complexity of the manifest variables by a 'quartimax' rotation, and seek that position that most nearly approximates the case where each variable has a nonzero loading on only one PC. These two rotations preserve the orthogonality of the PCs, which amounts to the assumption that they are all uncorrelated with one another. If, however, one's background assumptions allows that the PCs might or should be correlated, then other (nonrigid) rotations of the axes are possible. Fig. 2, for instance shows an 'oblimin' rotation.

Over the years, there have been a great deal of different kinds of rotations proposed in the literature (cf. Harman, 1976, chaps. 12–15 for an interesting discussion of the early history of this topic).

These rotational matters present only one issue that must be addressed. There are a great many more things to be settled before an empirically interpretable solution can be sought. E.g., a straightforward interpretation of a PCA assumes that the underlying influences are linearly related to the manifest data, and are not themselves internally structured in some important fashion. The latter would occur if a given axis was actually the result of some combination of disparate influences, which collectively had no relevant empirical interpretation. Similarly, attention to a PCA may reveal that seeking the 'best' axes, defined by the minimizing of the total squared deviations, is itself the wrong criterion to optimize. Some other type of optimization, such as an estimate of maximum likelihood, may be preferable. These kinds of follow-up analyses and methodological refinements illustrate how the explication-colligation loop is frequently quantitatively realized.

The explication of appropriate statistical conceptions is further made more difficult and important by the fact that PCA is only one of a growing body of techniques for reducing dimensionality and identifying latent variables. Indeed, it is not uncommon for researchers to analyze their data using multiple such methods, to see if different methods yield any interesting differences. For example, in factor analysis, the role of the variance of individual variables is effectively supplanted by an attempt to capture only the covariances of pairs of variables. Alternatively, some researchers want to strengthen the unassociated nature of the latent variables, so that they are not merely uncorrelated, but are statistically independent, i.e., $\Pr(X|Y) = \Pr(X)$. Recently, independent components analysis, a technique for attempting this, has been developed (Comon, 1994). The final technique I'll mention is the singular value decomposition (SVD), which can be viewed as follows. Suppose the original data set A is an $m \times n$ array, and suppose you wanted to find one m-dimensional vector $\mathbf{x}$ and one n-dimensional vector $\mathbf{y}$ such that the $m \times n$ matrix $A_1 = \mathbf{xy}^*$ provided the best approximation (in the sense of least squares) to A of any rank 1 matrix. SVD identifies those vectors. More generally, if you want to find the k ($\leqslant \min\{m, n\}$) pairs of vectors such that $A_k = \sum_{i=1}^{k} \mathbf{x}_i \mathbf{y}_i^*$ provided the best least-squares approximation to A of any rank k matrix, SVD would identify them.
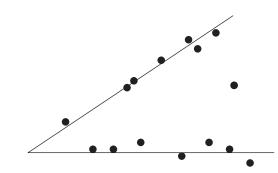

Fig. 2. An oblimin rotation, which obliquely rotates the axes, maximizing the extent to which individual data points are dependent upon exactly one axis.

Whewell's insistence upon the importance and difficulty of the ongoing process of explicating conceptions provides a lesson that some present day practitioners would do well to consider. To give an example, although SVD and PCA look rather different, mathematically speaking, they are quite similar, to the point that some researchers use the two terms interchangeably (e.g., Malinowski, 2002, p. 17).[5] But this can be unwise from a scientific perspective, and even more so from the vantage of the philosophy of science. Even in the brief characterizations given above, we can see that these techniques optimize different criteria, and thus extract different information from a data set. While they do often yield very similar results (i.e., these conceptions are quite similar from many practical perspectives), it is not uncommon for them to behave quite differently (e.g., Oblefias, Soriano, & Saloma, 2004; Conroy, Kolda, O'Leary, & O'Leary, 2000; Bell & Sejnowski, 1995, 1997; Phillips, Watson, Wynne, & Blinn, 2009). Failing to attend to these differences is tantamount to ignoring a unique kind of information that may reveal something important about the phenomena under study.

This last point can be generalized. As Whewell stressed, the explication of these various statistical conceptions of optimizing criteria, forms of optimal rotations, decompositions, etc. is very difficult and time-consuming. Moreover, contemporary statistical software allows for data to be thusly analyzed with just the press of a few buttons. Consequently, it is all too easy to misuse these techniques, by applying them to inappropriate data sets, misinterpreting the results, etc. As many methodologists have noted, this frequently occurs (e.g., Mauran, 1996; Armstrong, 1967, Fabrigar et al., 1999). Such overhasty practices occurred before the age of the computer as well. As Whewell notes, 'men often admire the deductive part of the proposition, the geometrical or algebraical demonstration, far more than that part in which the philosophical merit really resides' (II, p. 91). But, as he also noted, (II, p. 20), there is no logic of induction, no mechanical means by which the appropriate sort of analysis for a given data set can be calculated. The only thing that can be done is to engage in the difficult, excruciating, ongoing process of studying the various properties of these techniques, and continually refining them by explicating newer, better conceptions. When this is finally done well, we will have arrived at the proper conceptions, or even formal definitions of the latter. Thus, it seems that these statistical methods disagree with J.S. Mill, and side with Whewell, who insisted that definitions, if they are found at all, appear at the end of induction, not at the beginning. (cf. Snyder, 2006, pp. 108–110 for discussion). We don't engage in these 'hasty anticipations' because we can't. It is simply not possible to stare at 1269 vectors (each in R^{231}) and intuit with any credibility how many dimensions there really are, and where they lie.

It is worth observing that finding the right conceptions to bind together the facts is a central concern of many philosophically rel-

---

[5] Malinowski writes that since PCA and SVD 'produce essentially the same results, we use these terms interchangeably' (Malinowski, 2002, p. 17).

evant areas of contemporary cognitive science. E.g., many of the debates about the nature of concepts, mental representations, cognitive architecture, etc., ultimately concern the best way to organize the facts (e.g., Stich, 1992) A similar phenomena occurs in linguistics, where, as Whewell noted, the definitions of the technical terms come at the end, not the beginning, of inquiry (Johnson, 2007, §2).

As the discussion so far makes clear, statistical methods are generally restricted to discoveries of what Whewell called 'laws of phenomena', which concern 'the Order which the phenomena follow, Rules which they obey' (II, p. 95). They do not address the 'Powers by which these rules are determined, the Causes of which this order is the effect' (Ibid.) For Whewell, this would hold even for those theories that attempt to explicitly model causal structure in systems of equations (e.g., Pearl, 2000). Thus, talk of finding the 'true' axes should be understood in terms of finding the true statistical factors, which may suggest, but certainly do not determine, the true physical causes. In this sense of 'suggesting' causal factors, methods such as PCA partially resemble some principles in the philosophy of science, particularly Reichenbach's Common Cause principle (Reichenbach, 1956; cf. Artzenius, 2010). Roughly speaking, an (atemporal) version of this principle says that correlated phenomena share underlying common causes. A PCA does not unearth causes, but it does extract quantitative information about underlying correlational structure that can be relevant to the formation of quantitative causal hypotheses.

Although we often seek the right conceptions, as is well-known by users of these statistical methods,

> hypotheses may often be of service to science, when they involve a certain portion of incompleteness, and even of errour. The object of such inventions is to bind together facts which without them are loose and detached; and if they do this, they may lead the way to a perception of the true rule by which the phenomena are associated together, even if they themselves somewhat misstate the matter. The imagined arrangement enables us to contemplate, as a whole, a collection of special cases which perplex and overload our minds when they are considered in succession; and if our scheme has so much of truth in it as to conjoin what is really connected, we may afterwards duly correct or limit the mechanism of this connexion (II, p. 60; cf. the comparison with bookkeeping II, p. 81).

Indeed, the methods described above are often used simply to reduce the dimensionality of the data set, so as to be able to work with a more manageable number of variables. It is not uncommon to use a PCA or SVD with no intention of isolating the true latent structure, but only as an intermediary step in the search for other truths. For example, if the X_i variables in a regression equation $Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n$ are highly correlated, the confidence intervals for the estimations of b_is can become so large as to make any estimated values useless.[6] A common way of dealing with this problem is to perform the regression on the PCs, which are uncorrelated: $Y = b_0' + b_1' PC_1 + b_2' PC_2 + \ldots + b_k' PC_k$. Doing this can generate more reliable estimates, which can sometimes then help the researcher to better understand the relation between the Xs and Y (cf. e.g., Schott, 2005, pp. 97–99, 144–146 for discussion of the mathematical aspects of this).

Finally, we saw above that Whewell held that a successful explication-cum-colligation resulted in (merely) the discovery of a theory, which would later need to be confirmed (e.g., II, p. 51). This aspect of Whewell's philosophy is curious, and not always easy to understand (cf. Snyder, 1997, 2008). On the one hand, he claims that a good deal of work must go into the 'discovery' of a theory, and that this theory is the result of an inference. But on the other hand, he also claims that this inferential discovery is not the same as a confirmation. But it would seem that Whewell's form of discovery amounts to the formulation and *adoption* of a theory; that is, the discovery amounts to inferring that the theory is true. What more of confirmation is needed?

To understand Whewell's view, it is helpful to look at the same situation in statistics. We've seen that, although they are difficult and capable of yielding surprising conclusions, techniques like PCA and the like are not inferential methods. While a scientist may *choose* (as Romney and Indow did) to retain only a few dimensions from their analysis, this is not a statistical inference. The testing of dimensions for statistical significance, as mentioned above, is a matter for another set of techniques. Moreover, this latter stage can be important to the theory. E.g., as striking as Romney and Indow's findings are, by anyone's estimation, much more work needs to be done before their particular findings would be considered 'confirmed' E.g., it may be that, upon closer scrutiny, it will be found that, perhaps by the use of confirmatory methods—statistical significance tests for PCs, confirmatory factor analysis, etc—that an extra dimension needs to be retained above and beyond the three that Romney and Indow studied. And of course, there is nothing special about Romney and Indow's study; the careful scrutiny, and often adjustment, of a striking finding is a standard part of scientific inquiry. At the same time, in a Whewellian spirit, Roberts and Pashler (2000) criticize some of the scientific community for being overly lax. They cite a number of projects where parameter values for complex quantitative models were discovered, and the resulting model is given some degree of credence based on the resulting goodness of fit to the data assessed. But, as they note, and as Whewell would've noted, discovering, via inductive means, a good model (one that fits the facts well), is not the same as confirming it. After all, some models are so flexible that they can fit virtually any set of data, so the fact they do fit the data provides little reason to believe it is correct.

3.5. The Mill-Whewell debate

Interestingly, Whewell's verbal description of the mathematics of statistical reasoning is so accurate that it is even possible to reconstruct a famous objection to him, as well as (what I think is) a correct response on Whewell's behalf (found in e.g., Snyder, 2008, pp. 101–106).

In *A System of Logic* (Mill, 1949), Mill contends that Whewell's view of induction is not really induction at all. One of Mill's reasons for this criticism is that Mill believed that induction must be a form of ampliative *inference*. That is, the resulting theory cannot be a mere redescription of the facts; the theory must contain something new that was inferred from them. (E.g., inferring that 'All swans are white' goes beyond the comparatively small sample of observed swans, and extends the predicate 'white' to all swans, observed and otherwise.) Whewell, Mill held, included both genuine inductive inferences and mere redescriptions within the scope of (what Whewell called) induction. For example, Whewell considered Kepler's discovery of the elliptical nature of the orbit of Mars to be a paradigmatic case of induction. In contrast, Mill thought that, in using the equation for an ellipse, Kepler had merely found a convenient mathematical representation of the observed data.

---

[6] That is, suppose we want to predict the numerical value of Y from a number of variables Xs; e.g., we may want to predict GRE scores on the basis of GPA, SAT scores, parents' education level, and frequency of drug use. A natural first step would be to find a collection of weights for each of the Xs such that the resulting equation $Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n$ optimizes some criterion (such as least squares fit for the (n+1)-dimensional data set), and thus constitute a best estimate of the true values β_i. However, although the b_is might be optimal, they are only estimates of the β_is. If the Xs are highly correlated, the amount by which the β_is may plausibly deviate from these estimates can become so large as to render the estimates useless.

Replying on behalf of Whewell (and Kepler), Snyder (2008, p. 212) notes that from the point of view of the Earth, the observations would not appear elliptical. Thus, Kepler needed a theory of the Earth's motions to yield an elliptical interpretation of the data. Moreover, this particular ellipse appeared nearly circular, so Kepler needed to lean heavily on the theory's mathematical details in order to predict the proper ellipticality. Moreover, Kepler's theory goes beyond the data in two further ways that Whewell appreciated, but Mill apparently did not. First, Kepler applied the idea of an ellipse to the data. But since this idea was not one of the data points, the resulting theory does go beyond the data. Second, Kepler's theory was not limited to the observed data points, but went beyond them in predicting that all observations of Mars would fall on a certain curve. (cf. Snyder, 2008, pp. 211–214 for a detailed treatment of this topic, from which this discussion borrows heavily).

While Mill's argument is puzzling for many reasons, a present-day Millian might have an even easier time foisting the charge of mere redescription onto the methods discussed above. After all, redescription is exactly what these methods do. The decompositions we have examined are quite literally 'mere' redescriptions of the original data. A PCA, for example, is simply an orthogonal rotation (in Euclidean space) of the original coordinate axes determined by the measured variables. Viewed in the small, a PCA is the same kind of change as when the coordinates of a point in the plane change from [2, 3] to [3.536, .707] as we switch from the standard basis to the two (unit length) diagonals in the upper half-plane. (From the perspective of logic, this 'change' is analogous to switch the claims {P, Q, ∼R} for the logically equivalent set {P ↔ Q, ∼P → R, (Q ∧ ∼P) ∨ ∼R}.)

In a certain abstract sense, this charge is appropriate: A PCA is nothing more than a reorganization of the data along new bases. A PCA whose eigenvalues were nearly all equal in size would suggest that all the original dimensions should be retained, in which case the PCA would in fact be a mere redescription of the data. However, a crucial element of a PCA is the dimensionality reduction. When a PCA is performed, and, say only 3 of the 231 dimensions are retained, as in the Romney and Indow study, the scientist is inferring the theory that the remaining 228 dimensions capture only irrelevant noise in the data, and can therefore be safely disregarded. Similarly, this theory identifies a *very* specific 3-dimensional space where the data are located. Moreover, a PCA goes beyond the data in two further ways. First, as we've seen above, by using PCA, the scientist is implicitly claiming that certain extremal properties such as variance resolution and differential entropy are relevant (for some projects, they wouldn't be). So by using it, the scientist 'superinduces' some strong theoretical assumptions onto the data. Second, the results of a successful PCA are typically not limited to just the observations, but extend more broadly to the theoretical population. E.g., Romney and Indow's findings do not extend to just the 231 points in the visible spectrum (from 430 nm to 660 nm at 1 nm intervals), but apply to all the points in between the sampled points.

3.6. A final correct prediction

I briefly mention one more point of contact between Whewell's philosophy of science and contemporary statistics. Whewell admitted that the typical person will find the details of the workings of the sciences 'less pleasing' and 'neither so familiar nor so interesting' as many other topics (I, p. 13). For the typical person, these details 'will have in them nothing to engage his fancy, or to warm his heart' (I, p. 14). Moreover, Whewell's own meta-scientific study is 'abstruse and uninviting', filled with 'the most dark and entangled questions', so that the ordinary reader will find the project 'obscure or repulsive' (I, p. 13). Having taught introductory sta-

tistics for several years, I regret to say that these sentiments have been non-quantitatively realized in contemporary times.

4. Conclusion

In this paper, we have seen a genuinely impressive level of detail at which Whewell's Discoverer's Induction quite simply gets it right about statistics. Moreover, we saw that success in the statistical case is strong evidence for general correctness of the view. To this end, I can only mention that there are an enormous number of details in Whewell's work, above and beyond what I have discussed, that map neatly onto contemporary statistical-cum-empirical practice. The most glaring exception is perhaps Whewell's theologically based belief that our fundamental ideas are the right ones.

Finally, we can note that one of Whewell's leading contemporary interpreters writes that '[Whewell's] philosophy of science … is … a view worthy of our attention today' (Snyder, 1997, p. 601). In this paper, we've seen ample reason to enthusiastically endorse this claim.

Acknowledgements

References

Armstrong, J. S. (1967). Derivation of theory by means of factor analysis or Tom Swift and his electric factor analysis machine'. *The American Statistician, 21*, 17–21.

Artzenius, F. (2010). Reichenbach's common cause principle. *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/physics-Rpcc/> Accessed 20.02.11.

Basilevsky, A. (1994). *Statistical factor analysis and related methods*. New York: Wiley-Interscience.

Bell, A. J., & Sejnowski, T. J. (1995). An information–maximization approach to blind separation and blind deconvolution. *Neural Computation, 7*, 1129–1159.

Bell, A. J., & Sejnowski, T. J. (1997). The 'independent components' of natural scenes are edge filters. *Vision Research, 37*, 3327–3338.

Buchdahl, G. (1991). Deductivist versus inductivist approaches in the philosophy of science as illustrated by some controversies between Whewell and Mill. In M. Fisch & S. Schaffer (Eds.), *William Whewell: A composite portrait* (pp. 311–344). Oxford: Oxford University Press.

Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing, 36*, 287–314.

Conroy, J. M., Kolda, T. G., O'Leary, D. P., & O'Leary, T. J. (2000). Chromosome identification using hidden Markov models: Comparison with neural networks, singular value decomposition, principal components analysis, and Fisher discriminant analysis. *Laboratory Investigation, 80*, 1629–1641.

Desrosières, A. (1998). *The politics of large numbers: A history of statistical reasoning*. Cambridge: Harvard University Press.

Eckart, C., & Young, G. (1939). A principal axis transformation for non-Hermitian matrices. *Bulletin of the American Mathematical Society, 45*, 118–121.

Efron, B. (1986). Why isn't everyone a Bayesian? *The American Statistician, 40*, 1–5.

Fabrigar, L., MacCullum, R., Wegener, D., & Strahan, E. (1999). Evaluating the use of exploratory factor analysis in psychological research'. *Psychological Methods, 4*, 272–299.

Fisch, M. (1985). Whewell's consilience of inductions – An evaluation. *Philosophy of Science, 52*, 239–255.

Fisch, M. (1991). A philosopher's coming of age: A study in erotetic intellectual history. In M. Fisch & S. Schaffer (Eds.), *William Whewell: A composite portrait* (pp. 31–86). Oxford: Oxford University Press.

Forster, M. (1988). Unification, explanation, and the composition of causes in Newtonian mechanics. *Studies in History and Philosophy of Science, 19*, 55–101.

Hald, A. (2007). *A history of parametric statistical inference from Bernoulli to Fisher, 1713–1935*. New York: Springer.

Harman, H. (1976). *Modern factor analysis* (3rd ed.). Chicago: University of Chicago Press.

Horn, R. A., & Johnson, C. R. (1985). *Matrix analysis*. Cambridge: Cambridge University Press.

Johnson, K. (2007). The legacy of methodological dualism. *Mind and Language, 22*, 366–401.

Krantz, David H., Duncan Luce, R., Suppes, Patrick, & Tversky, Amos (1971). *Foundations of measurement volume 1: Additive and polynomial representations.* San Diego: Academic Press, Inc..

Malinowski, E. R. (2002). *Factor analysis in chemistry* (3rd ed.). New York: Wiley-Interscience.

Mauran, M. D. (1996). Metaphor taken as math: Indeterminacy in the factor analysis model. *Multivariate Behavioral Research, 31*, 517–538.

Mill, J. S. (1949). *A system of logic.* London: Longmans, Green and Co. (First published 1843).

Oblefias, W. R., Soriano, M. N., & Saloma, C. A. (2004). SVD vs PCA: comparison of performance in an imaging spectrometer. *Science Diliman, 16*, 74–78.

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge: Cambridge University Press.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, 2*, 559–572.

Phillips, R. D., Watson, L. T., Wynne, R. H., & Blinn, C. E. (2009). Feature reduction using a singular value decomposition for the iterative guided spectral class rejection hybrid classifier. *ISPRS Journal of Photogrammetry and Remote Sensing, 64*, 107–116.

Reichenbach, H. (1956). *The direction of time.* Berkeley: University of Los Angeles Press.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107*, 358–367.

Romney, A. K., & Indow, T. (2003). Munsell reflectance spectra represented in three-dimensional euclidean space. *Color Research and Application, 28*, 182–196.

Ruse, M. (1991). William Whewell: Omniscientist. In M. Fisch & S. Schaffer (Eds.), *William Whewell: A composite portrait* (pp. 87–116). Oxford: Oxford University Press.

Schott, J. R. (2005). *Matrix analysis for statistics.* Hoboken: Wiley Interscience.

Snyder, L. J. (1997a). Discoverer's induction. *Philosophy of Science, 64*, 580–604.

Snyder, L. J. (1997b). The Mill-Whewell debate: Much ado about induction. *Perspectives on Science, 5*, 159–198.

Snyder, L. J. (2006). *Reforming philosophy.* Chicago: University of Chicago Press.

Snyder, L. J. (2008). 'The whole box of tools': William Whewell and the logic of induction. In D. M. Gabbay & J. Woods (Eds.), *Handbook of the history of logic, volume 4* (pp. 165–230). The Netherlands: Elsevier.

Stewart, G. W. (1993). On the early history of the singular value decomposition. *SIAM Review, 35*, 551–566.

Stich, S. P. (1992). What is a theory of mental representation? *Mind, 101*, 243–261.

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900.* Cambridge: Belknap Press.

Stigler, S. M. (1990). *Statistics on the table: The history of statistical concepts and methods.* Cambridge: Harvard University Press.

Stuart, A., & Ord, K. (1994). *Kendall's advanced theory of statistics: Volume I: Distribution theory.* London: Hodder Arnold.

Whewell, W. (1819). *An elementary treatise on mechanics.* Cambridge: J. Deighton and Sons. <http://www.archive.org/details/elementarytreati00whew> Accessed 20.02.11.

Whewell, W. (1825). A general method of calculating the angles made by any plane of crystals. *Philosophical Transactions of the Royal Society of London, 115*, 87–130.

Whewell, W. (1833). *Astronomy and general physics, considered with reference to natural theology.* Bridgewater Treatise III. London: William Pickering. <http://www.archive.org/details/astronomygeneral00whew> Accessed 20.02.11.

Whewell, W. (1836). Researches on the tides – 6th series. On the results of an extensive system of tide observations made on the coasts of Europe and America in June 1835. *Philosophical Transactions of the Royal Society of London, 126*, 289–341.

Whewell, W. (1838). *The doctrine of limits with its applications namely conic sections, the first three sections of Newton, the differential calculus.* Cambridge: J. and J. J. Deighton. <http://www.archive.org/details/doctrinelimitsw00whewgoog> Accessed 20.02.11.

Whewell, W. (1847). *The philosophy of the inductive sciences* (2nd ed.). New York: Johnson Reprint Corporation.

Whewell, W. (1856). Mathematical exposition of certain doctrines of political economy, third memoir. *Transaction of the Cambridge Philosophical Society, 9*, 1–7.

Whewell, W. (1858). *History of the inductive sciences, from the earliest to the present time* (3rd ed.). London: J. W. Parker. <http://www.archive.org/details/historyinductiv00whewgoog> Accessed 20.02.11.

Wilson, M. (2006). *Wandering significance.* Oxford: Clarendon.

Yeo, R. (1993). *Defining science: William Whewell, natural knowledge, and public debate in early Victorian Britain.* Cambridge: Cambridge University Press.