NOÛS 00:0 (2014) 1–27
doi: 10.1111/nous.12075

*Realism and Uncertainty of Unobservable Common Causes in Factor Analysis**

KENT JOHNSON
Department of Logic and Philosophy of Science UC Irvine

Abstract

Famously, scientific theories are underdetermined by their evidence. This occurs in the factor analytic model (FA), which is often used to connect concrete data (e.g. test scores) to hypothetical notions (e.g. intelligence). After introducing FA, three general topics are addressed. (i) Underdetermination: the precise reasons why FA is underdetermined illuminates various claims about underdetermination, abduction, and theoretical terms. (ii) Uncertainties: FA helps distinguish at least four kinds of uncertainties. The prevailing practice, often encoded in statistical software, is to ignore the most difficult kinds, which are essential to FA's underdetermination. (iii) What to do: some suggestions for dealing with these hardest types of uncertainty are offered.

1. Introduction

A perennial topic in the philosophy of science concerns the underdetermination of a theory by its evidence. The idea is familiar: for any given body of evidence, there will always be multiple theories that capture it in some appropriately equivalent manner. Thus, with respect to the given evidence, these theories are equivalent, and hence it is underdetermined which (if any) of them should be favored as "correct". Quine, in a famous paper on the topic, offers a representative statement of the issue:

> Under-determination lurks where there are two irreconcilable formulations each of which implies exactly the desired set of observation conditionals plus extraneous theoretical matter, and where no formulation affords a tighter fit. [....]
>
> Here, evidently, is the nature of under-determination. There is some infinite lot of observation conditionals that we want to capture in a finite formulation. Because of the complexity of the assortment, we cannot produce a finite formulation that would be equivalent merely to their infinite conjunction. Any finite formulation that will imply them is going to have to imply also some trumped-up matter, or stuffing, whose only service is to round out the formulation. There is some freedom of choice of stuffing,

*I am grateful to an anonymous reviewer for a great deal of helpful criticism. The audiences at UC Merced and the UC Boulder Conference on Measurement across the Sciences also offered many helpful comments. This paper benefited from further feedback from: Jeff Barrett, Rick Dale, Alistair Isaac, Brian Skyrms, and Jeff Yoshimi.

and such is the under-determination. [Quine, 1975, 324] (Cf. [Quine, 1951, 1955, v. O. Quine, 1972] for similar sentiments; Laudan [1990] carefully examines Quine's views on underdetermination.)

Although underdetermination is often discussed in very general terms [Kukla 1996], this paper examines it in the specific context of the statistical technique of common factor analysis, here after dubbed the *FA model*, or FA for short. (This model is the key component of family of techniques known as exploratory factor analysis.) FA's roots trace back to Charles Spearman's work on intelligence testing [Spearman, 1904, 1922, 1927]. Suppose e.g. 3000 students take an exam with 100 questions involving mathematics, reading comprehension, spatial reasoning, etc. The students' answers are recorded, so that there are 3000 observations on the 100 measured (or "manifest") variables. One might wonder whether there are some underlying unmeasured "latent" abilities that are responsible for the students' performances on the exam. If so, then how many such abilities are involved? Which test questions draw from which abilities to what degrees? How are these abilities related to one another? In terms of individual differences, which students possess more of a given ability than which, and by how much? And for a given question, do similar students recruit the same abilities to the same degree? What do the data have to say about all this?

There's an enormous amount to be said about the particular issues of human intelligence; cf. [Hunt, 2011, Sternberg and Kaufman, 2011]. However, intelligence testing only illustrates FA, which attempts to bridge the gap between observable items of limited interest (e.g. test scores), and unobserved hypothetical entities of much greater interest (e.g. human abilities). FA might also be used if the data came from measuring the sensitivity of 3000 retinal cone cells of monarch butterflies to 100 different wavelengths of the electromagnetic spectrum in order to learn about the number of kinds of retinal cones, their relations to one another and to various light wavelengths, etc. Or the data might've come from measuring 3000 ocean regions in 100 different ways in order to try to develop some hypotheses about what (unobserved) causes were operating where and to what degree. Or we might be examining 100 chemical analyses from 3000 tissue samples from a newly discovered rodent, to learn about their (unobserved) diets, hereditary characteristics, etc. ; cf. e.g., [Basilevsky, 1994, Jolliffe, 2010, Brereton, 2003] for many more examples. ("Latent" variables aren't restricted to hypothetical unseen forces or magnitudes: astronomers and biologists use them to assess the "shape" of complex objects like nebulae and lobster claws.)

Over the years, FA has received virtually no attention in the philosophical literature (although cf. [Baird, 1987, Glymour, 1998]). This is a shame, because FA possesses many philosophically interesting features centered around its underdetermination. Moreover, this underdetermination can be precisely characterized and analyzed. But FA is not a mere toy for philosophical rumination; instead, it is of considerable actual practical use, present within a diverse array of broader empirical theories. It is also a flexible and general model that can be adjusted to suit various circumstances, both philosophical and empirical. Not only are there literally thousands of actual theories built upon it; but countless more can easily be imagined for philosophical purposes. Furthermore, although I do not discuss this here, the issues below also apply to various types of currently popular causal models; e.g. [Woodward, 2003, Spirtes et al., 2000, Pearl, 2000].

So why hasn't hasn't FA gotten more attention? There are at least two main reasons. First, factor analysis is a notoriously difficult technique, reducing some expositors to profanity [Gould, 1996, 268].[1] Fortunately, many of the central philosophical issues can be brought to light with a relatively small investment of technical material. Second, FA has a rather mixed reputation; e.g., [Armstrong, 1967, Fabrigar et al., 1999]. Infamously, many studies have simply misused FA and the results produced therein. This, however, isn't a problem with the tool but with its misapplications. The present discussion, if anything, helps clarify how FA should be used.

FA supports many more theses than I can argue for here. However, this paper argues for the following: (i) FA supplies a philosophically interesting type of underdetermination. In fact, it can clarify several key issues. For example, (ii) when assessing whether a putative case of underdetermination holds for "all possible observations", several matters must be carefully distinguished; when they are, the nature of underdetermination and the difference between theory and evidence become much less clear. (iii) Several determinate approaches (e.g. regression analysis, principal components) are mathematically but not conceptually similar to FA, precisely because they lack the latter's extra "randomness". This point is easily missed, even by practicing scientists. (iv) The amount of underdetermination in an FA study can be measured, in terms of the correlations between the various potential key theoretical terms. This assessment isn't the end of the story, even for FA; however, it does introduce to the philosophical discussion of underdetermination several considerations which deserve more widespread attention. (v) In contrast to the received wisdom, FA is not a "brutally empirical" technique. Instead, it is a particularly strong and clear case of abductive inference—one where the capacity to make he data likely is actually sacrificed in favor of explanatory desiderata, particularly Reichenbach's Principle of the Common Cause. (vi) FA helps to distinguish four kinds of uncertainties that may be present in a scientific theory. Underdetermination is associated with the two most intractable forms, which are routinely ignored, often as an institutionalized practice. (vii) Finally, three recommendations are briefly scouted for more explicitly dealing with FA's underdetermination.

This paper is organized as follows. §2 introduces FA with a simple example. §3 argues that factor indeterminacy maps nicely onto the more philosophical notion of underdetermination. This requires a sharp distinction between FA and regression analysis—philosophically speaking, the former's theoretical terms must, but the latter's cannot, "go beyond the data" in a certain precise sense. §4 further explores several philosophical issues concerning FA's underdetermination. §5 considers what types of uncertainties are present in FA, and examines the general practice of ignoring them. §6 contains three speculative suggestions for how to better address this underdetermination. §7 concludes the body of the paper. Several technical appendices are included.

2. A Simple Introduction to the FA Model and Its Underdetermination

This section introduces some of the main ideas behind the FA model by way of an artificially small example. The key ideas, however, generalize to all standard applications within the factor analysis literature. The model is more formally characterized below in Appendix A; cf. e.g. [Basilevsky, 1994, Mulaik, 2010, Bartholomew et al., 2011] for further technical discussion.

Suppose that a great many subjects take a test consisting of just four items x_1, x_2, x_3, x_4. The raw data of these subjects' scores might look something like that given in the table below in (1):

Table 1. Hypothetical raw test scores

(1)

	Meg	Sue	Bob	Ann	Fred	Ned	...
x_1	94	78	69	88	92	79	...
x_2	89	76	72	88	96	78	...
x_3	91	81	78	59	91	83	...
x_4	90	78	70	91	90	93	...

(Realistically, the number of test items would be much larger; however, this does not affect the points to be made here.) Suppose also that the four test items are found to be positively correlated with one another: for any two test items, subjects who did well on one of them tended to do well on the other, and similarly for those who did poorly on one of them. Suppose the correlations between x_i and x_j, denoted r_{ij}, are: $r_{12} = .48$, $r_{13} = .40$, $r_{14} = .32$, $r_{23} = .30$, $r_{24} = .24$, $r_{34} = .20$. At this point, our target empirical question is whether there are a few underlying abilities that the subjects possess to varying degrees, which are in turn indicated to varying degrees by the various tests. The goal behind FA is to use the the raw data in (1) to construct/identify some statistical candidates which are optimal in a certain sense, and which might plausibly represent such abilities.

We can think of FA as being structured around two fundamental ideas. The first is that FA realizes Hans Reichenbach's Principle of the Common Cause; [Reichenbach 1956]; cf. also [Sober, 1984, Uffink, 1999, Pearl, 2000, Suppes and Zanotti, 1981]. This principle says that when two or more phenomena are probabilistically dependent, this dependency should be due to some shared causal structure affecting them both. In a slogan, there should be "no correlation without causation" [Pearl, 2000, 61]. FA satisfies Reichenbach's principle by treating the correlations amongst $\{x_1, x_2, x_3, x_4\}$ as the primary explananda, and constructing a small set of variables $f_1, ..., f_q$ that collectively account for them, statistically speaking.[2] In particular, the primary task is to render all the x_i mutually uncorrelated, conditional on $f_1, ..., f_q$. For present purposes, Reichenbach's Principle can be formally defined as the conjunction of (FAii–iii) in Appendix A. (When there are multiple fs, we might also require either that they be uncorrelated, or that there are higher-level common factors accounting for their correlation; this detail will not matter below.)

The second key idea is to construct these fs using some ideas familiar from ordinary regression analysis. In particular, we would like to find a comparatively small number of fs such that linear combinations of them predict, in some sense, the various x_i:

$$
\begin{aligned}
x_1 &= \lambda_{11} f_1 + \lambda_{12} f_2 + \ldots + \lambda_{1q} f_q + e_1 \\
x_2 &= \lambda_{21} f_1 + \lambda_{22} f_2 + \ldots + \lambda_{2q} f_q + e_2 \\
x_3 &= \lambda_{31} f_1 + \lambda_{32} f_2 + \ldots + \lambda_{3q} f_q + e_3 \\
x_4 &= \lambda_{41} f_1 + \lambda_{42} f_2 + \ldots + \lambda_{4q} f_q + e_4
\end{aligned} \tag{2}
$$

Here, the e_i are the *specific factors*—those statistical parts of the x_i that are not represented by the fs, which are the *common factors*. Since the f are typically of primary theoretical importance, I follow the convention of referring to them simply as the "factors". The λs are the factor *loadings*; they are the fixed numbers that relate the latent factors to the manifest data.

In practice, q is much smaller than k, so if (2) holds at all, it does so as a substantial empirical fact. For our example, we can suppose that, as an empirical fact about the data in (1), the correlations in question can be captured by only one common factor: $q = 1$. Thus, suppose that, leaving aside issues of sampling error, (3) holds (with $f_1 = f$): [3]

$$
\begin{aligned}
x_1 &= .8f + e_1 \\
x_2 &= .6f + e_2 \\
x_3 &= .5f + e_3 \\
x_4 &= .4f + e_4
\end{aligned} \tag{3}
$$

As part of (3) fitting the FA model, we can assume that Reichenbach's principle is satisfied, in the sense that all pairs of the latent variables $\{f, e_1, e_2, e_3, e_4\}$ are uncorrelated (at least up to issues of sampling error).

The intuitive idea here is that the latent factor f that FA uncovers might represent some empirical property (or ability, etc.) of the population from which the subjects were sampled. In practice, this task is often aided by considering the factor loadings, i.e., the coefficients .8, .6, .5, .4 above. Is there some unobserved property of the subjects that would naturally be seen to be quite influential on the first test item, somewhat less so on the fourth item, etc? If so—or if the data are strong enough to motivate adjusting background theory, etc. to better fit these results—then the usual further study of f and what it might represent can continue. f, then, would be our statistical characterization of the unobserved property we wish to study. (There is, of course, no guarantee that the factor loadings will support any such natural interpretation of f.)

For example, the four xs might've been four distinct problems involving the mental rotation of objects or scenarios. Although this shared demand may not have been obvious to the researcher, by uncovering the statistical pattern in (3), FA can suggest such an empirical interpretation. Moreover, the factor loadings can suggest how much rotational ability each x requires of the subjects. (I focus on a simple case of just one factor and a ready interpretation; in practice matters are often much more complicated.)

Obviously, when more than 4 xs are involved, these matters become much more reliant on methods like FA. Also, as with any other theory, the resulting characterization of f, the loadings, etc. are subject to further exploration and testing. E.g., significance tests of given sets of factor loadings is at the heart of "confirmatory factor analysis".

The procedure just described has been the basis of over a century of research. Leaving aside a huge amount of detail, we can focus on what a latent variable like f might tell us about about some (potential) unobserved empirical property of interest. This matter would be much easier to address if there was only one such f that fit (3) (and the other assumptions of FA; cf. Appendix A). However, f is not unique.

Early in the history of factor analysis, the mathematician Edwin B. Wilson noticed that the FA model does not determine f [Wilson, 1928a, 1928b, Lovie and Lovie, 1995]. In general, even when the loading coefficients are held fixed, there are still infinitely many possible solutions to any set of common and unique latent factors. Over time, many authors have developed and explored this phenomenon; e.g. [Piaggio, 1931, 1933, Ledermann, 1938, Kestelman, 1952, Guttman, 1955, Heermann, 1964, 1966, McDonald, 1974, Schönemann, 1971, Maraun, 1996, Bartholomew, 1981, 1996, Mulaik and McDonald, 1978]. For useful historical overviews of this topic; cf. [Steiger and Schönemann, 1978, Lovie and Lovie, 1995] and [Mulaik, 2010, chap. 10].

It would be easy to get lost in the great many details concerning even the narrow issue of the underdetermination of the single latent common factor f. However, the following informal characterization should suffice; cf. Appendix B for details.

Following standard notation, let $\hat{f}$ be the ordinary least-squares regression of any candidate f onto the xs:

$$\hat{f} = b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4, \tag{4}$$

for some fixed bs. (This might seem rather odd, since regression analysis typically requires some values of f, and that is precisely what is not available. However, it is a mathematical fact that in this context, $\hat{f}$ exists, and is unique and calculable.) To construct a candidate f, though, it is necessary to augment $\hat{f}$ with another variable u. u is required because $\hat{f}$ by itself doesn't have enough variance to act as a latent common factor f by itself. Suppose, as is common, that the xs and f are stipulated/normalized to each have a variance of 1. In that case, the variance of $\hat{f}$ will be less than 1, and so will need to be augmented. But (4), it turns out, also captures the desired relationship between any candidate latent variable f and the given manifest xs. So it will also be necessary that u be uncorrelated with the xs. Thus, every candidate f has the form:

$$f = \hat{f} + u, \tag{5}$$

where u is any variable uncorrelated with the xs, and such that f has a variance of 1 (cf. Appendix B). Thus, there will always be infinitely many equivalent but different candidates for this key theoretical term.

This, then, is the underdetermination of FA: f, the latent structure that is often of primary interest, necessarily contains an "extra" random part. Moreover, there are infinitely many such choices—indeed, since the variance of u can always be scaled to fit, the only serious requirement is that u must have nothing to do with the empirical evidence!

For example, in the simple case of (3), the loadings (.8, .6, .5, .4) collectively yield a variance of $\hat{f}$ of only about .74, so an extra part u with a variance of .26 will have to be added to it to construct the latent variable f. Thus, to construct an f, we can select *any* variable whatsoever independent from the xs, and scale it to fit.[4] E.g., u could have the familiar form of a normal distribution, as in (6a). However, it could also have, say, a highly skewed distribution like that in (6b).[5] And of course, these are only two options: u could be multimodal, discrete, a mixture of various distributions, etc.

(6)

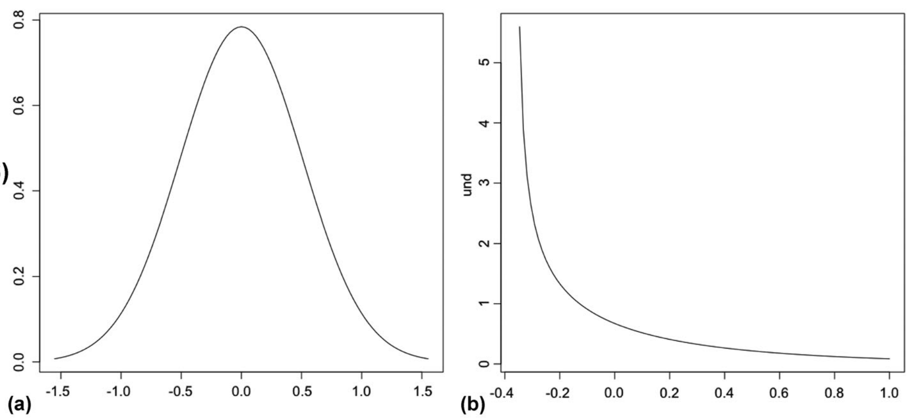

Figure 1. Two possible underdetermined parts, u^a and u^b

Call the distribution in (6a) u^a, and the one in (6b) u^b. Similarly let $f^a = \hat{f} + u^a$ and $f^b = \hat{f} + u^b$ be the resulting latent common factors; likewise with e_i^a, e_i^b, etc. (The unlabeled variables (f, u, etc.) are used below for arbitrary common factors, underdetermined parts, etc.)

f^a and f^b are quite distinct, but as latent common factors, they satisfy (3) equally well, yielding identical values for the xs. They thus are a precise, unambiguous, and concrete case of underdetermination of a theoretical entity by the empirical evidence of the xs, and the strong theoretical assumptions of FA. Empirically speaking, if the common factor represents, say, a given human ability, then f^a suggests that this ability is symmetrically distributed. f^b, however, suggests that there is a small population of individuals with an exceptionally high amount of the ability, and no similarly sized population of exceptionally impoverished individuals.[6]

Often, much of the purpose of a factor analytic study ultimately rests on the numeric values of the latent variables. These values correspond to empirical questions of fundamental importance, such as "How intelligent is individual i"?[7] That is, we often care about the manifest data in the xs and their intercorrelations because of what they reveal about the latent factors; and the latter is often of interest because of what they say about particular individuals (persons, blood samples, oceanic regions, etc.). These most basic quantities are the *factor scores*—the particular values that particular individuals have on the latent factor. They are central to what follows.

3. Underdetermination: Philosophical Aspects

This section discusses FA's underdetermination in more detail. Quine's description, quoted above, is a useful starting point. Does FA's underdetermination fit it?

Consider the simple example just given. Quine writes of "two irreconcilable formulations" each implying the same empirical statements.[8] This is exactly what we have with f^a and f^b: they (along with the e^as and e^bs) each imply exactly the same values of the xs. Thus, they are equally good common factors. They only differ in their "extra" bit, the u^a, u^b, which is the "trumped-up matter, or stuffing, whose only service is to round out the formulation" of f^a and f^b (Ibid.). But with these extra bits, the resulting latent factors f^a, f^b, etc. constitute the "alternative hypothetical substructures" at the core of the equivalent theories [Quine, 1975, 313].

Importantly, not only are f^a and f^b equivalent rivals; in accordance with Quine, "no formulation affords a tighter fit" [Quine, 1975, 324]. That is, no common factor f better fits the empirical data. But this raises an important question. As noted, f^a and f^b share $\hat{f}$, the best (least-squares) representation of the common factor in terms of our evidence (the xs). So why isn't $\hat{f}$ alone preferable to f^a or f^b? After all, $\hat{f}$ differs from the latter exactly by omitting the arbitrary part, the Quinean "stuffing" that rounded out the common factors. Doesn't $\hat{f}$ afford a tighter fit?

Surprisingly, perhaps, the answer is unambiguously no. $\hat{f}$ doesn't suffice because it is not "random" enough. This might seem odd, since the purpose of any factor f is to account for the xs. But more can be said: Thus, let $f^{\#} = a_1x_1 + \cdots + a_4x_4$ be *any* linear combination of xs whatsoever. Could $f^{\#}$ be a common factor, as in (3)? No:

(7) **Proposition.** FA implies that $f^{\#}$ is not a common factor for (3).

Because of its importance, Appendix C provides an explicit algebraic proof that uses only very minimal assumptions.

(7) answers our question. Only the various fs ($= \hat{f} + u$, for some u) provide Quine's desired "best fit", because their shared "best fitting" statistical component $\hat{f}$ is not a possible candidate. That is, $\hat{f}$ (but not f) is the best fit in terms of least-squares regression; however the fs (but not $\hat{f}$) are the best, and only, fit in terms of FA.

(7) can also be understood in terms of Reichenbach's principle. If f statistically accounts for the correlations in the xs, the remaining behavior in the es is uncorrelated. Thus, each e requires its own dimension, or degree of freedom. But since f is also uncorrelated with them, it too needs its own dimension/degree of freedom. So we need at least five such dimensions, but the xs only supply four. Thus, Reichenbach's principle always requires at least one more dimension than the empirical data offer, and that is precisely what produces the underdetermination in question. Regression approaches only use the dimensions that the data supply, and so they cannot respect Reichenbach's principle. Thus, $\hat{f}$ literally doesn't supply enough dimensions to do the job of a common factor. It is in this sense that a model that uses regression doesn't contain enough randomness. (7), of course, applies to any linear combination of the xs; thus similar conclusions can be drawn about a variety of other approaches, including principal component analysis, FA's most well-known alternative; Jolliffe [2010]. In sum, we have

(8) In general, with a linear model, you can satisfy Reichenbach's principle, or you can have determinacy, *but you can't have both.*

(The linear model takes the form of (2), where the fs are to be determined. Some background mathematical provisos are needed in various cases, but the point here is that (8) holds quite generally.) As (7) shows, if you accept indeterminacy, there is no easy way out that involves accepting only those solutions that are most favored by the data. Instead, all the solutions are equally (dis)favored. In fact, the situation is especially dire for those wishing to avoid underdetermination. As Louis Guttman's theorem shows (cf. Appendix B), underdetermination only requires an exceedingly weak form of Reichenbach's principle. In particular, it only requires that the fs be uncorrelated with the es; the es needn't be uncorrelated, but merely linearly independent. The fs can be correlated amongst themselves, and some further correlations may remain in the e. Intuitively, the part of Reichenbach's principle that really matters for underdetermination here is that screening off the common causes doesn't affect the remaining correlational structure in the es that hasn't yet been accounted for.

The ideas just discussed are fundamental to FA. However, they are frequently missed, even by practicing psychometricians. A recent instance of this occurs in [Beauducel 2013]; the details are in a footnote.[9]

3.1. "All Possible Observations"

So far, we've seen that FA's underdetermination fits neatly with Quine's classic philosophical statement. However, philosophers often require that the underdetermination should persist across "all possible" observations. The idea is that the underdetermination should run deep—it shouldn't occur merely because, e.g., a large enough sample hadn't been collected, even though it could've been. Prima facie, this requirement seems straightforward; in truth, matters are much more subtle.

To begin, note that as a general type of statistical model, FA represents literally thousands of highly varied actual examples. Some of them contain further relevant assumptions beyond FA; many don't. But the various features of underdetermination will always be highly dependent on the particular details of the case at hand. Thus, we can only ask whether particular applications of the FA model are underdetermined for all possible data. However, the following considerations hold in general.

In multivariate statistics, new data rarely arrive as a single number. Instead data appear as new individuals or new variables (or both); cf. (1).[10] But these are two very different kinds of additions: one could add new individuals measured on the original variables, or one could add new variables on which the original individuals are measured. (This distinction is of course not unique to statistical theorizing.) I take these two in turn.

Increasing the sample size by adding individuals can reduce the sampling error associated with, e.g., the estimation the λ coefficients, etc. However it is irrelevant to FA's underdetermination, which holds at the population level. Thus, in one very natural sense, the philosophical issue has an affirmative answer: FA's underdetermination holds across all possible observations, simply because it is a population-level phenomenon, not a sample-level one. Of course, as with any empirical theory, there might be observations that undermine the whole theory; the present point, though, is that if the theory is true, no such observations can affect the underdetermination (between say f^a, f^b, etc.).

I now turn to the much more interesting matter of adding new variables to a factor-analytic study. Importantly, these few remarks hardly scratch the surface of this large and important issue.

Consider the general case of underdetermined theories, as in e.g. van Fraassen's well-known discussion of the universe's absolute motion within a Newtonian theory; [van Fraasen, 1980, 44ff.]. The thrust of this example is that, *by itself*, Newton's physical theory of mechanics and gravitation is consistent with any claim that the universe has a constant absolute motion of r, for any non-negative value of r. Now of course, such underdetermination might not hold if Newton's theory were *augmented* with, say, the claim that the universe has an absolute motion of 0, or that St. Peter, who is always to be believed, appeared and declared as much, etc. This is only to say that a theory's underdetermination depends entirely on its content: strengthen the theory appropriately, and the underdetermination is thereby lessened. In other words, even if underdetermination must persist as new *observations* are added, it needn't persist as new *bits of theory* are added.

Obviously, St. Peter may not be available to alleviate Newtonian underdetermination. Likewise, there is no guarantee that new variables are available to eliminate FA's underdetermination. Moreover, simply assuming that there will always be such variables to be found doesn't reflect the realities that, e.g., psychometricians face when trying to define and understand personality traits, cognitive abilities, and the like. (This sentiment is not new; e.g., Lipton argues that "[w]hat counts for our actual epistemic situation is not ideal underdetermination by all possible evidence, but the much greater actual underdetermination by the evidence we now have" [Lipton, 2004, 200].)

Even more troubling is the fact that, when adding new variables, the very distinction between theory and observations mentioned above becomes much less clear. After all, data like that in (1) also partly serve to identify f, the key theoretical term of interest. (Of course, one may also have some background theoretical constraints on the empirical phenomena that f itself is thought to represent; but the original point still applies to whatever unknown aspects there are that motivated the study in the first place.) Suppose e.g. you add a new variable x_5 to (3). In such a case, you have thereby switched to a different, albeit related, FA model that employs five manifest variables instead of four. Let f^5 be the new latent factor thus extracted from the five variables. Statistically speaking, f^5 is distinct from the original f (or fs—I here set aside the fact that f is underdetermined). But the larger question is; can f and f^5 be said to both represent the same empirical phenomenon? To the extent that they can, x_5 is a bit more evidence regarding the nature of the phenomenon; to the extent that they cannot, f^5 is part of a new theory about a somewhat different phenomenon, a potential rival to the theory f. However, it commonly happens that the only thing known about these two "extents" are that they're both very limited: f^5 is kinda-sorta pointing at the same thing as f, but not exactly. At this point, the theory-evidence distinction is truly a Quinean will-o-the-wisp.

For example, we earlier supposed that $x_1, \ldots, x_4$ were cognitive tasks involving mental rotation. Suppose that x_5 involves self-location in a scene ("You walk north of the church, which is east of the school. You turn left, and walk.... Are you facing the river?"). Assume that f is held to represent an ability for mental rotation. Does f^5 as well? Or does it represent a more expansive ability for spatial reasoning? There doesn't seem to be much of an answer here. Moreover, if no variables or other considerations are forthcoming, there may never be; indeed, there may be no "fact of the matter" how this aspect of our cognitive abilities is carved up.[11]

In sum, the strict requirement that underdetermination hold for "all possible observations" is met in the relevant sense for many applications—actual and philosophical—of the FA model. However, this fact is of lesser importance than the complexities relating theory and evidence encountered in actual practice.

4. Further Issues

The issues discussed so far bear on several other topics in the philosophy of science. This section addresses four of them.

4.1. Robustness and Invariance

FA's underdetermination is a robust phenomenon, in at least two respects. First, its mathematical basis is very general. FA's underdetermination does not involve any particular statistical machinery (involving testing procedures, estimators, etc.).

Rather it is largely independent of all such considerations. Similarly, it makes very little use of probability theory, as the proofs of the various results mentioned above show. Indeed, the most important role of probability theory is to justify the Euclidean structure of a vector space.[12] At heart, FA's underdetermination is fundamentally algebraic (or geometric, if you prefer). In short, FA's underdetermination doesn't depend on a wide variety of specialized high-level assumptions. Rather, it lurks deep within the bones of the mathematical underpinnings of much empirical research. Thus, its frequent occurrence and profound resistance to elimination is unsurprising.

Second, FA's underdetermination is also robust in terms of parametric invariance. To return to our example, suppose you switch from the FA structure that uses f^a to the one that uses f^b. Then, in addition to identical reconstructions of the manifest data, the following theoretical features will also remain invariant: (i) the particular *numbers* of latent fs and es involved; (ii) the *means and standard deviations* of every x, f, and e; (iii) every *correlation* between any two distinct variables; i.e., all pairs of $\{x_1, ..., x_k, f_1, ..., f_q, e_1, ..., e_k\}$; and (iv) every *loading coefficient* λ_{ij}. Thus, the underdetermination between f^a and f^b extends far beyond a mere equivalent reconstruction of the data—it also includes a great deal of theoretical structure, regardless of how the latter was obtained. In contrast, many other "observationally equivalent" approaches do not share these invariances. E.g., Thomson's bonds model radically violates all of (i)–(iv); [Thomson, 1916; Bartholomew et al., 2009a]. Similarly, the rotational/scaling issues developed by Thurstone radically violate at least (iii) and (iv); [Gould, 1996; Thurstone 1947]. (Notice that even in our simple example, (iii) alone involves 30 theoretical constraints; a realistic case would involve a great many more.)

4.2. Is Underdetermination Real?

Recently, some philosophers have suggested that we don't really encounter the pairs of equivalent theories that underdetermination requires. E.g., John Norton argues that such theories differ only in their Quinean extra "stuffing", and that we should:

> treat all such superfluous structure as representing nothing physical.... For we have two theories with a common core fully capable of returning all observations without the additional structures in question. Moreover ... the additional structure may lack determinate values. For example, as stressed by Einstein famously in 1905, the ether state of rest of Lorentz's theory could be any inertial state of motion. Because of the perfectly symmetrical entry of all inertial states of the motion in the observational consequences of Lorentz's theory, no observation can give the slightest preference to one inertial state over another. So its disposition is usually understood not to be an unknowable truth but a fiction. [Norton, 2008, 36]

Norton's description fits FA, but but the conclusion does not follow. After all, all candidate fs share a "common core [$\hat{f}$] fully capable of returning all observations [the xs] without the additional structures [u] in question". Moreover, there's an important sense in which the u "lack determinate values", since they can be constructed from nearly any probability distribution(s). And no xs "can give the slightest preference" to one such u over another. But from all this, it does not follow that u and hence f is "but a fiction". Instead, u is essential to FA – giving up this extra stuffing requires denying at least one of the empirical claims of the model, which may be a factual mistake. (Moreover, as we've seen above with f^a vs. f^b, different fs have different theoretical consequences.) Of course, u is also irrelevant to the point-predictions that $\hat{f}$ alone can make about observations. This suggests that mere predictive capacity *cannot*, as a mathematical fact, be the only desideratum in a use of FA. If it were, then FA might be used to obtain the estimated regression weights, with the regression $\hat{f}$ used by itself in an ordinary regression model.[13] This matter resurfaces repeatedly below.

4.3. How Different Are Equivalent Theories?

Since we can't simply dismiss underdetermination as never occurring, it's natural to wonder "how" underdetermined a given theory is—is it enough to be worried about? Norton, e.g., suggests probably not: "If we are to be able to demonstrate observational equivalence of the two theories, the theoretical structures of the two theories are most likely very similar. While it is possible that they are radically different, if that were the case, we would most likely be unable to demonstrate the observational equivalence of the two theories" [Norton, 2008, 34–35].[14]

Leaving aside the "most likely" proviso (although cf. §5), FA offers an oddly specific response to the claim that its equivalent theories are "radically different". FA's key theoretical components are the fs; how different might they be? In our simple model that uses just one latent factor, a natural measure of the maximum possible difference between equivalent candidates is the minimum possible correlation between any f, f^* pair. Interestingly, this can be determined. If R is the multiple correlation between any latent f and the xs, then the correlation between f and a minimally correlated counterpart f^* is:

$$2R^2 - 1 \tag{9}$$

Cf. [Guttman, 1955, 73].[15] That is, for any common factor f, there exists another one f^* such that the correlation between them is $2R^2 - 1$. Thus, there is no centroid factor f^{**} offering some kind of compromise, being moderately correlated with all of them.[16]

Suppose, for example, that Norton's criterion for being "radically different" is that two factors f and f^* should be uncorrelated. By the above this occurs iff $R \leq .707$. But this condition is frequently met in actual research. E.g., a great many correlations in the literature on human intelligence lie in the .3 to .6 range; if the relevant R is too, there will be equally appropriate latent factors that are *negatively* correlated with one another, to the tune of $-.28$ to $-.82$! Increasing values of f would then predict, perhaps strongly, decreasing values of f^*.[17] In the example in (3), the multiple correlation between f and $\mathbf{x}$ is .86, and so the value of (9) is .48. Thus, only ($.48^2 =$) 23% of the variance of one such latent factor can be predicted from the other.

In short, FA's underdetermination offers a concrete case where there is much more to be said than simply that the rival theories are not all that different after all.

Obviously, this feature of does not automatically extend to other cases of underdetermination. Moreover §6 argues that (9) shouldn't be understood as exhausting even FA's underdetermination. However, it does present a goal that we might aim for even in less clear cases.

4.4. Abductive Inference

In this section, I wish to argue that the contrast between FA and regression displays how FA encodes a strategy of abduction, or "inference to the best explanation" (IBE). Of course, both regression and FA involve theoretical constructions that are "best fits" of the manifest data, so they both possess some rather superficial abductive credentials. Conversely, both regression and FA outcomes are frequently subjected to further inferential testing, so neither represents the full inferential story; these additional details don't affect the present discussion, however.

The key idea behind IBE is that scientific inferences are frequently driven by *explanatory* considerations, and not just the ability to render the data likely (or at least comparatively so); e.g. [Lipton, 2004, 2008]. It's always desirable for a theory to render the data more likely than its rivals do, but according to IBE, a theory should do so in a way that increases our understanding of the relevant phenomena (to the extent the theory is true).

On the one hand, note that by including u, f has a larger variance than $\hat{f}$; but since u is uncorrelated with the xs, the u only adds extra "noise" in terms of f's ability to predict the xs. That is, u spreads f's distribution out more widely, orthogonally to the xs. Thus, the data (i.e. the xs) are less probable when f is used instead of $\hat{f}$, and so the likelihood is decreased: $p[x|f] < p[x|\hat{f}]$, where p is the appropriate probability (density) function.[18] On the other hand, this reduction in likelihood comes with an increase in the explanatory capacities of the theory. As we've already seen, the factors (f) of FA satisfy Reichenbach's Principle of the Common Cause, whereas their regression estimates ($\hat{f}$) do not. Thus, to the extent that, as is currently popular in the philosophical literature, Reichenbach's principle is regarded as an important explanatory desideratum, factors like f provide something of value that their more predictive counterparts $\hat{f}$ don't.

We've just seen that FA doesn't merely value certain explanatory virtues, it *pays* for them in likelihood, the only fungible epistemic currency here. Thus, FA supplies a precise instance of Peter Lipton's comment that "[e]ven when our main interest is in accurate prediction or effective control, it is a striking feature of our inferential practice that we often make an 'explanatory detour'" [Lipton, 2004, 65].[19] (Indeed, the size of this "detour" is measurable.) In this sense, the received wisdom is exactly wrong in viewing factor analysis, in Stephen Jay Gould's words, as a "brutally empirical technique" [Gould, 1996, 346]. A truly brutal empiricism would emphasize predictability at all costs; [Sober, 2008, Cartwright, 1983, Kukla and Walmsley, 2004, Leplin, 2004].[20] As we've seen, regression approaches do just this, albeit at the cost of the the explanatory powers of FA—i.e., FA might be used merely in order to estimate $\hat{f}$, and then summarily ignored in favor of $\hat{f}$ and its predictions. This moderately inconsistent brutally empirical approach has in some ways become an institutionalized practice; cf. §5.1.

(I suspect that Gould calls FA "brutally empirical" because its Reichenbachian assumptions are very general and apply to an extremely broad range of types of empirical inquiry. Thus, it's natural to take Reichenbach's principle for granted, in contrast to other desiderata that apply much more narrowly; e.g., that the data should have a lognormal distribution, or that the most significant f should account for nearly all of the data, or that it should be twice the size of the second one, with no other significant fs, etc.)

5. Uncertainties and FA

What sorts of uncertainty are created by FA's underdetermination? Does it supply anything *new*? Initially, it might seem that it doesn't. After all, statistical models are already awash in uncertainty. Does underdetermination merely supply more of the same?

A proper understanding of FA requires us to distinguish at least four different kinds of uncertainty, labeled (U1)–(U4) below. (By stipulation, "uncertainty" here is a very general notion, and need not be based on a probability distribution.) The short story is this: (U1) concerns noise that masks a signal. (U2) applies when the signal itself contains a "known unknown", and (U3) applies when the signal contains an "unknown unknown". (U4) is like (U3) except that the signal contains unknown, possibly varying, "unknown unknowns". Of these forms of uncertainty, the most important distinction is the separation of (U1)–(U2) and (U3)–(U4). Unfortunately, (U3)–(U4) are rarely even acknowledged by statisticians and practitioners—with some reason, I suggest below. Let us turn, then, to these uncertainties.

(U1): Uncertainty from external sources. This is probably the most common sort of uncertainty; it includes many forms of sampling error. E.g., a simple regression model represents the theoretical relationship of interest between x and y as $E[y|x] = \alpha + \beta x$, but the model admits extraneous noise: $y = \alpha + \beta x + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$. I mention (U1) only to set it aside, since it doesn't bear directly on the present issues of underdetermination.[21]

(U2): Uncertainty from a known constituent random variable. (U2) is the type of uncertainty supplied by a random *constituent* of the statistical phenomena of interest. For example, consider a signal-detection model, where the goal is to detect, on each trial, whether only noise (ϵ) has occurred, or whether the signal has occurred along with the noise ($s + \epsilon$), where $\epsilon \sim N(0, 1)$, say. If s is constant, e.g., $s = .4$, then only (U1)-type uncertainty is present. But when s has its own distribution, e.g., $s \sim N(.4, .7)$, this represents (U2)-type uncertainty. (U2) is closer to the role u plays, since it is part of f, and not simply external noise. (U2) is the uncertainty about f's value that remains in particular instances, *given* fixed, known xs and f. But selecting one such f over its rivals is precisely what is at issue, so (U2) is a far cry from the uncertainty inherent in FA's underdetermination.

(U3): Uncertainty from an unknown constituent random variable. According to (U3), our Quinean stuffing has a fixed but unknown distribution. It is the remaining uncertainty about f's value, given a fixed, known x and a fixed but *unknown* f.

Unlike (U2), there's little to say about (U3) (e.g., is u symmetric? Skewed? Multimodal? Discrete?) (U3) is much closer to the uncertainty that underdetermination creates. However, even (U3) should be augmented.

(U4): Uncertainty from unknown multiple realizations. (U4) is like (U3) except that it removes the epistemic stability brought by the assumption that the unknown constituent has a single fixed distribution. To see this, imagine constructing an FA model for groups A and B (or one group at different times, etc.). The data in x may suggest that only one model, and thus a single f (or vector $[f_1, ..., f_q]'$) is involved. However, the key theoretical term(s) f may be different for the different groups, even though x does not suggest this. Indeed, even a project whose very purpose is to explore whether or not two groups involve different fs might run afoul of (U4). Such a study could result in identical estimates for the distribution of the xs, as well as identical estimates of the λs, f, and the es, up to sampling error. But by (U4), the core of underdetermination, a conclusion of no difference could still be incorrect (cf. the discussion of (6) above). In such a case, switching across groups results in an undetectable change to a different FA model. This would happen if the kinds of evidence the researchers have access to is given by x, but the various groups' fs have different distributional properties that are statistically independent of it.[22] Thus, although the available evidence cannot show it, this is a form of potential "multiple realization" of different fs [Fodor, 1974, 1975]. (Of course, this multiple realization doesn't occur at the physical level, though presumably physical differences undergird the distributional differences.)

5.1. Coping Mechanisms

Unsurprisingly, over time there have been several responses to the phenomenon of factor underdetermination; e.g., [Mulaik and McDonald, 1978, Bartholomew, 1981, Maraun, 1996]. By far, the dominant approach for decades has been complete disregard. E.g., Steiger and Shonemann write: "If a single striking fact dominates the history of factor indeterminacy, it is the tendency of the psychometric community to ignore the problem and its implications" [Steiger and Schönemann, 1978, 170]. Similarly, "modern researchers routinely compute factor scores. These same researchers appear to be completely unaware of factor score indeterminacy. Consequently, factor scores are derived and left unevaluated, and their potentially adverse effects on the results of subsequent analyses are ignored" [Grice, 2001, 432].[23]

This pattern of disregard raises two question: how and why does it happen? Both questions have interesting answers; I take them in turn.

Factor scores are commonly estimated by one of two very similar methods, due to Thomson and Bartlett respectively; e.g., Bartholomew et al. write "Only two practically useful sets of scores have ever been proposed: those of Thomson and Bartlett. In practice they are either identical or very close" [Bartholomew et al., 2009b, 581]; [Bollen, 1989, 305]. Since they are linearly related (e.g., in (3) they yield identical z scores), for our purposes, they needn't be distinguished. Both methods are essentially regressions of the common factors onto the xs (i.e., $\hat{f}$), which we saw in the previous section is incompatible with FA. Moreover, "[f]rom a practical point of view the question of how to actually calculate factor scores has not been taken any further forward" [Bartholomew et al., 2009b, 570]. (This is not quite true; e.g., [Grice, 2001, ten Berge, et al. 1999] discuss other point estimators. However, since they are linear combinations of the xs, by (7) they face corresponding difficulties.) From this perspective, all uncertainty is mere sampling error (U1). So to whatever extent the uncertainties of underdetermination, (U3) and (U4), are acknowledged (typically not at all), they are treated as if they were of the most benign, most familiar, sort. This curious attitude is intentional: e.g., "[Bartlett] treats the specific factors [sc. the es] as *random errors*.... [which] brought the estimation problem within the ambit of standard Fisherian inference" [Bartholomew et al., 2009b, 577, emphasis added]. Intentionally or not, this practice completely sidesteps underdetermination: e.g. Thomson "wanted to minimize the error he made whatever individual he picked, providing that their scores are determined *by the same model*" [Bartholomew et al., 2009b, 577, emphasis added]. However, we've seen in (U4) that this assumption of "the same model" is precisely what we *cannot* assume. E.g., for the concrete example(s) above with four manifest variables, which model should we assume: one that uses the symmetric f^a, or one that used the skewed f^b? Or a different one altogether? Of course, there may be theoretical reasons for eliminating some distributions in certain empirical applications; but nothing suggests all the relevant alternatives can always be eliminated in the various empirical applications of interest. Similarly, the issue is not with the point estimates themselves; e.g, $\hat{f}$ is an unbiased estimator, but so is the constant estimate 0, for that matter.

Nothing promotes disregard like automation. Bartlett and Thomson estimations are the ones used by the popular statistical software R with the *factanal* command [R Core Team 2013]; cf. also the options in the sophisticated *psych* package [Revelle 2013]. Moreover, there are no standard methods for accounting for the uncertainty of these point-estimates wrought by the uncertainty of the nature of f, in the sense of (U3)–(U4). Indeed, the best advice seems to be based largely on hope and faith: "Thus researchers should refrain from *too fine* comparisons of standings on factor scores, for they may be asking more from the factor score estimates than they can provide" [Bollen, 1989, 305–6, emphasis added].

Thomson, Bartlett et al. had good practical reasons for their proposals; but such strategies are still highly imperfect. This of course raises our second question of *why* there would be such disregard, to which I now turn.

The principal reason why (U3)–(U4) have received virtually no attention by statisticians and practitioners, I suspect, is that they are extremely hard to manage. This is especially clear in relation to the much more tractable difficulties presented by (U1)–(U2), which may thereby attract more attention; cf. the discussion of Bartlett above. After all, an explicit quantitative grip on either (U3) or (U4) requires some theoretically principled way to countenance essentially *every possible random variable* with a finite second moment, and there's no obvious way to do this. Indeed, this seems to be the core of the difficulties that (U3)–(U4) present: their uncertainties concern an uncountable infinitude of distributions which collectively lack the kind of structure that might support a theoretically satisfying organization, by a higher-level probability distribution or some other means. Thus, the problem is not merely that there are empirically equivalent rival theories, it's also that the total set of them is largely unstructured. Of course, one can always *impose* some organization on this set, but this isn't the problem. The problem isn't that there isn't any way to organize all the underdetermined distributions, it's that there are far too many such ways, and no good means for favoring some over others. Thus, the challenge is to find a way to do so that is not question-begging, but is more empirically useful than simply assigning random numeric labels to the distributions.

Could a Bayesian perspective help, whereby a prior probability distribution is placed on the various distributions that the us might realize? If so, then the uncertainty of (U3) could be brought into the general fold of probability theory, using the ordinary tools of Bayesian statistics. However, this move doesn't seem very promising, primarily because of the extreme lack of constraints on the possible distributions u (and hence f) can take. Any such prior distribution on the us would be a distribution over almost all possible probability distributions, and it is not clear how this could be done in a non-question-begging way. One pragmatic move that is common in Bayesian statistics is to use an "improper" distribution, whereby e.g. some constant value, say 1, is assigned to every possible prior value (each possible distribution of u in the present case). Such a prior measure assigns an infinite value to the total prior set, and hence is not a probability distribution. However, such a prior assignment can nonetheless often combine with the rest of the statistical machinery to produce satisfactory posterior distributions; e.g. [O'Hagan and Forster, 2004, 74–77]. However, this tactic doesn't characterize the uncertainty that underdetermination presents; it simply avoids it, in the hopes that doing so won't affect later results. (Corresponding remarks hold for various other mathematical tricks, like compressing unbounded continua into finite intervals and inducing uniform distributions on the latter.) Additionally, by representing the total uncertainty in (U3) as infinitely large, it is thereby incomparable with other forms of uncertainty that are measured with standard probability distributions.

6. What to Do About Underdetermination?

This final section briefly offers three speculations for dealing with FA's underdetermination. In general, it seems that when underdetermination is robustly present in theories that matter, the former should constitute an object of study, not a reason for hopelessness, as some have suggested; e.g. Maraun [1996].

Minimum correlation is only part of the story. Unless the minimum correlation between candidate factors (cf. (9)) is very high, the available fs will differ in terms of many of their various properties. E.g., two individual variables might be moderately correlated but have importantly different shapes—one might be multimodal with various peaks and valleys, while the other could be discontinuous, perhaps possessing discrete point masses at particular locations. When fortune smiles, background theory may eliminate some of these possibilities, but there's no guarantee. In general, there are *lots* of ways for moderately correlated variables to differ. Moreover, some of these differences may matter empirically. E.g., if a latent factor is supposed to measure some human ability, it could be very important to know whether its distribution is bimodal, with a substantial valley separating the "haves" from the "have-nots", or unimodal with a fat right tail, indicating a higher proportion of exceptionally able persons. But attending only to how well correlated the various fs are can easily obscure these possibilities.

Chebyshev's Inequality. Many of the usual measures of uncertainty for a given value of a particular f require knowing the latter's distribution. Consider, e.g. a confidence interval for the mean of a normally distributed x with a known variance of σ^2. The fact that $P[|x-\mu| \leq 1.96\sigma] = .95$ is then used to produce the familiar endpoints $x \pm 1.96\sigma$. This strategy, which might help with (U1)–(U2), suggests a way to deal with the very different case of (U3). Chebyshev's inequality states that $P[|x-\mu| \geq c\sigma] \leq 1/c^2$. This inequality only requires that x has a finite variance, which is the one thing that is known about the us. Thus, even when the distribution is unknown, it can still be said that there is at least a 95 % chance that a value of u will be within 4.47σ of μ. Such boundaries provide a range of values where the true value plausibly lies, assuming no "error" of the familiar (U1) sort. If sampling error is also present, as it surely is, these boundaries do not reflect it, and should be widened accordingly.

In some ways, Chebyshev's inequality is a reasonable compromise. On the one hand, it provides a sharp boundary that prevents an "anything goes" attitude toward the uncertainty in question. Interestingly, since all candidate latent factors have the same variance, this boundary remains invariant across them all. Thus, the boundary applies even with regard to (U4), which does not assume a single fixed unknown u (or f). I.e., even if the distribution of f is constantly changing in ways we're unaware of, the variance doesn't change, so neither will any information derived from Chebyshev's inequality. On the other hand, the uncertainty of point estimates is considerably increased over, say, what would be provided by an ordinary confidence interval alone.[24]) It's no criticism of a practice to say that it incorporates some uncertainties that others ignore. Nevertheless, it does amount to accepting that genuinely compelling results from a factor analysis may be harder to obtain than is typically acknowledged.

Underdetermination as an additive component. Finally, we will often want to aggregate multiple kinds of uncertainty into a global assessment of the total uncertainty present in a given estimate or collection of estimates. Suppose for the moment that we have figured out how to do this for various individual kinds, including (U3) and (U4), say. The remaining challenge is how to combine them in some informative manner. In the absence of a mathematical model that characterizes how to do this, it seems reasonable to to simply add them together. The idea is to extend, by analogy, the fact that the variance of a sum of uncorrelated random variables is simply the sum of the variances. Thus, the variance of any *particular* f is: $Var[\mathrm{f}] = Var[\hat{f}] + Var[\mathrm{u}]$, but if one wants to also take account of the fact that u is *not* determined, which is characteristic of (U4), it seems reasonable to simply add this further uncertainty to whatever else is being tallied up. Intuitively, this treats "u" as a kind of higher-order random variable, since one is accommodating the variation in its various possible distributions—each of which have their own variance (the same in each case). In any case, without a probability distribution on the entire class of probability distributions that u can take, treating higher-order variation this way does not, unlike the lower-order summing of variances, follow from probability theory alone.

7. Conclusion

In many ways, FA raises more questions than it addresses, but in a good way. In particular, it exchanges (parts of) a few vague and general queries with a bevy of much more precise statements, whose answers are sometimes mathematically tractable. E.g., we saw above that for FA, the general issue of whether theories are (very) underdetermined can be partly addressed exactly by (9), but that some very real issues remain, not about theories in general, but about probability distributions. Moreover, these latter issues lie at the heart of the really hard problem with FA's underdetermination: the quantification over the infinitude of highly varied probability distributions. Similarly, issues of empiricism vs. rationalism are transformed, via considerations of predictiveness, likelihood, and the criteria for "best" explanations, into considerations of regression, FA, and the order vs. rank of a matrix or random vector. Some of these more specific issues, I've argued at length, have no obvious solutions. But their intractability doesn't lessen their precision; the former difficulties remain whether we are exact or vague about what's at issue. At the very least, turning a general issue into several precise but difficult ones clarifies where matters stand.

Appendix A: Formal Definitions and Properties of the FA Model

This section presents some key formal features of FA; cf. e.g. [Basilevsky, 1994, Mulaik, 2010, Bartholomew et al., 2011] for more discussion. Throughout, individual random variables are denoted by lowercase letters, (x, u, f, e,...); random vectors are finite sequences of random variables, given in boldface ($\mathbf{x}$, $\mathbf{u}$, $\mathbf{f}$, $\mathbf{e}$, ...); matrices of real numbers are denoted by uppercase letters (B, Σ, Λ, Φ,...); transpose is denoted with a hash ($\mathbf{x}'$, Λ', ...). A background assumption is that:

(10) Every individual variable (x, f, e, u, etc.) has a mean of 0 and a positive finite variance: i.e, for all such variables y: $E[y] = 0$, and $0 < E[y^2] < \infty$.

The basic form of FA is:

$$\text{(FA)}\quad \begin{array}{ll} \text{(i)} & \mathbf{x} = \Lambda\mathbf{f} + \mathbf{e} \\ \text{(ii)} & E[\mathbf{fe}'] = \mathbf{0} \\ \text{(iii)} & E[\mathbf{ee}'] = diag[e_1^2, ...e_k^2] = \Psi^2. \end{array}$$

$\mathbf{x}$ is a $k \times 1$ vector of *manifest* (aka observed, measured) variables, Λ is a $k \times q$ matrix of fixed constants (the factor *loadings*), $\mathbf{f}$ is a $q \times 1$ vector of unobserved latent variables (the *(common) factors*), and $\mathbf{e}$ is a $k \times 1$ vector of latent variables (the *specific factors* of the manifest $\mathbf{x}$).

The "Fundamental Theorem of Factor Analysis", which illustrates Reichenbach's principle, follows immediately from FA. The theorem observes that the covariance matrix Σ of the manifest variables $\mathbf{x}$ can decomposed into two parts:

$$\Sigma = \Lambda\Phi\Lambda' + \Psi^2, \tag{11}$$

where $\Phi = E[\mathbf{ff}']$ is the covariance matrix of the common factors $\mathbf{f}$. The proof of (11) is immediate: $\Sigma = E[\mathbf{xx}'] = E[(\Lambda\mathbf{f} + \mathbf{e})(\Lambda\mathbf{f} + \mathbf{e})'] = \Lambda E[\mathbf{ff}']\Lambda' + \Lambda E[\mathbf{fe}'] + \Lambda' E[\mathbf{ef}'] + E[\mathbf{ee}'] = \Lambda\Phi\Lambda' + \Lambda\mathbf{0} + \Lambda'\mathbf{0} + \Psi^2 = \Lambda\Phi\Lambda' + \Psi^2$. QED. Only (FAi) and (FAii) were used to derive (11) But by (FAiii), Ψ^2 is diagonal, and so all the correlational structure between distinct manifest variables is completely determined by $\Lambda\Phi\Lambda'$.

Although unnecessary, a few additional simplifying assumptions are often made:

$$E[f_i f_j] = 0, \text{ for every } f_i,\ f_j \text{ in } \mathbf{f}\ (i \neq j). \tag{12}$$

$$E[x_i^2] = 1, \quad \text{for every } x_i \text{ in } \mathbf{x}\ . \tag{13}$$

$$E[f_i^2] = 1, \quad \text{for every } f_i \text{ in } \mathbf{f}\ . \tag{14}$$

With these extra assumptions, $\Phi = \mathbf{I}$, and $\Sigma = R$, the correlation matrix for $\mathbf{x}$. Thus, (11) reduces to $R = \Lambda\Lambda' + \Psi^2$, and all correlational structure is given by the factor loadings: $r_{ij} = \lambda_i\lambda_j'$, where λ_i is the ith row of Λ (and similarly for j). So if there is only one latent variable ($q = 1$), Λ is simply a $k \times 1$ column, and so r_{ij} will simply be the product of the loadings of the single factor on x_i and x_j. Thus, for all a, b, c, and d in $\mathbf{x}$, $r_{ab}r_{cd} = \lambda_a\lambda_b\lambda_c\lambda_d = \lambda_a\lambda_d\lambda_b\lambda_c = r_{ad}r_{bc}$. Thus, $r_{ab}r_{cd} - r_{ad}r_{bc} = 0$, which is Spearman's "tetrad differences".

Appendix B: The Underdetermination of FA

Below is a simplified characterization (in updated notation) of Louis Guttman's Theorem 2. Let $\hat{\mathbf{f}} = B\mathbf{x}$ and $\hat{\mathbf{e}} = C\mathbf{x}$ be the (ordinary least squares) regressions of $\mathbf{f}$ and $\mathbf{e}$ onto $\mathbf{x}$. Then:

(15) **Underdetermination Theorem** [Guttman, 1955, 70–72]: $\mathbf{x} = \Lambda\mathbf{f} + \mathbf{e}$, and $E[\mathbf{fe}'] = 0$

if and only if there exists a random vector $\mathbf{u}$ such that:

(i)	$\mathbf{f} = \hat{\mathbf{f}} + \mathbf{u},\ \mathbf{e} = \hat{\mathbf{e}} - \Lambda\mathbf{u}$
(ii)	$E[\mathbf{uu}'] = \Phi - E[\hat{\mathbf{f}}\hat{\mathbf{f}}']$
(iii)	$E[\mathbf{xu}'] = 0$

As above, it is assumed that there are k variables in each of $\mathbf{x}$ and $\mathbf{e}$ and $q > 0$ variables in $\mathbf{f}$. All other vectors, matrices, etc. will be assumed to be of the appropriate sizes. In short, (15) establishes that if there exists one latent random vector **f** that satisfies FA, then there are infinitely many. Moreover, if the correlation matrix for **x** can be factored as in (11), then by Guttman's Theorem 1 [Guttman, 1955, 69], there does exist an **f**.

Appendix C: Proof of (7)

This section proves a generalization of (7), where there are k manifest variables, and q latent common factors.

(16) **Lemma 1.** If (FAi) holds, and $\mathbf{f} = B\mathbf{x}$ for some B, then the set of latent variables has a rank of at most k, and so is not linearly independent.

Proof. If $\mathbf{x} = \Lambda\mathbf{f} + \mathbf{e}$, and $\mathbf{f} = B\mathbf{x}$, then $\mathbf{e} = (I - \Lambda B)\mathbf{x}$. Setting $\mathbf{h} = [\mathbf{f}'\mathbf{e}']'$ and $C = [B'(I - \Lambda B)']'$, we have $\mathbf{h} = C\mathbf{x}$. Thus, each of the $q + k$ latent variables $\{f_1, ..., f_q, u_1, ..., u_k\}$ is a linear combination of the k manifest variables, and so they all lie within the span of the k manifest variables. Thus the span of **h** has dimension k or less, and so the the set of latent variables is linearly dependent. □

(17) **Lemma 2.** If (FAi, ii, iii) hold, and **h** is of rank less than $k + q$, then the linear dependencies lie exclusively within **f**.

Proof. Suppose the span of $\{f_1, ..., f_q, e_1, ..., e_k\}$ is less than $k + q$. Then at least one of these variables, h, is a linear combination of some of the remaining ones:

(18) $$h = a_1 h_1 + \cdots + a_j h_j,$$

where every a is nonzero.

First, suppose that h is one of the specific factors, e. Suppose that some $h_{i'}$ that composes e is also a specific factor, e_i. Then by (FAiii), $E[ee_i] = 0$. But by (FAii) and (FAiii), e_i is also uncorrelated with all the other factors composing e: $E[h_{i''}e_i] = 0$, for all $h_{i''}$ *except* when $h_{i''} = e_i$. Thus, $E[ee_i] = E[(a_1h_1 + \cdots + a_jh_j)e_i] = a_1E[h_1e_i] + \cdots a_{i'}E[e_ie_i] + a_jE[h_je_i] = a_{i'}E[e_ie_i]$ By (10), $E[e_ie_i] = \sigma^2 > 0$. Thus, $E[ee_i] = a_{i'}\sigma^2 \neq 0$, a contradiction. Suppose instead that some $h_{i'}$ that composes e is a common factor, f_i. By similar reasoning, we have again both that $E[ef_i] = 0$ and that $E[ef_i] \neq 0$. Thus, h in (18) cannot be a specific factor.

Suppose that h is a common factor f, and that some $h_{i'}$ that composes e is a specific factor, e_i. Again using similar reasoning: $E[fe_i] = 0$ and $E[fe_i] = a_{i'}\sigma^2 \neq 0$. Thus, h in (18) is a linear combination of only common factors. (In such a case, some of the common factors are correlated, and Φ is of less than full rank.) □

(7) **Proposition.** FA implies $\mathbf{f} \neq B\mathbf{x}$, for all B.

Proof. Suppose not. By lemma 1, **h** is linearly dependent, and has rank at most k. But by the argument of lemma 2, **e** has rank k. so the $q > 1$ variables in **f** are dependent on **e**, which contradicts lemma 2. □

Notes

[1] Indeed, Gould focuses on a much simpler technique, principal component analysis, which does not produce the underdetermination in question.

[2] If the xs are not multivariate normal, Reichenbach's Principle might be realized only for correlations, without achieving the stronger notion of statistical independence. I do not address this detail here.

[3] In practice, the estimation of factor loadings is a significant part of factor analysis. However, in the imagined case involving plenty of subjects, this is not too difficult. E.g., a sample of 3000 observations (with data generated randomly from the model in (3)) yielded the estimate $[.79, .59, .52, .41]'$.

[4] I.e., pick any variable u^0 with finite, nonzero variance σ^2, and set $u = \frac{\sqrt{.26}}{\sigma} \times u^0$.

[5] In (6a), $u\tilde{}N(0, .26)$, and in (6b), $(u + .36)\tilde{}\Gamma(.5, 1.39)$. In the latter distribution, .5 provides a substantial skew, and along with 1.39 yields a variance of .26. Translation by .36 ensures a mean of 0.

[6] Notice also that f^a and f^b are univariate, and are not linearly related. Thus, this form of indeterminacy is very different from, and much stronger than, the rotational issues popularized by Gould [1996]; cf. [Thurstone, 1934, 1938, 1947]. These two phenomena are often conflated; e.g. [Kasper and Ünlü, 2013, 3, 4, 18]. For a careful treatment, cf. [Heermann, 1964, 1966]. In particular, the rotational issues only concern the choice of axes for a *given* subspace. The present issue renders underdetermined which *subspace* to base the empirical data in. That is, f^a and f^b are axes for two distinct 1-dimensional subspaces (of the underlying infinite-dimensional subspace of random variables).

[7] Or, how intelligent is i along some given dimension of interest, if a multi-faceted view of intelligence is adopted.

[8] Quine's actual statement is made in terms of "observation conditionals" which are fully contextualized "standing" statements, whose antecedents are the standing background conditions that enable the entire conditional to be mathematically derivable from the theory in question.

[9] Beauducel writes that "[$\hat{\mathbf{f}}$ is] compatible with the defining equation of the common factor model" [Beauducel, 2013, 291], which directly contradicts (7). Beauducel's argument is based on his Lemma 1, which states that

($*$) If $\mathbf{x} = \Lambda\hat{\mathbf{f}} + \mathbf{e}$ and $\hat{\mathbf{f}} = B'\mathbf{x}$, then $B = \Sigma^{-1}\Lambda\Phi$,

where Λ, Σ, and Φ are as defined in Appendix A; cf. also [Beauducel, 2013, 290]. B is of course the correct regression matrix for the FA model. However, it doesn't follow that (*) is a sound derivation of it. It is easy to construct counterexamples where, e.g. $\Lambda \neq \mathbf{0}$, and $\Phi \neq \mathbf{0}$, and $\mathbf{e} = \mathbf{x}$. With little effort, it can be shown that $B = \mathbf{0} \neq \Sigma^{-1}\Lambda\Phi$. Obviously, further assumptions beyond those stated are needed.

Beauducel's proof starts by postmultiplying the first equation in (*) by $\mathbf{x}'$. This yields: $\Sigma = \Lambda B\Sigma + E[\mathbf{ex}']$. However, *pace* Beauducel, we cannot assume that $E[\mathbf{ex}'] = \Psi^2$, since the argument assumes only (FAi) and not (FAii, iii) (which is what Beauducel uses on p. 290). Leaving this aside, the purported derivation of B' continues:

$$B = (\Lambda'\Lambda)^{-1}\Lambda'(\Sigma - \Psi^2)\Sigma^{-1} = (\Lambda'\Lambda)^{-1}\Lambda'(\Lambda\Phi\Lambda')\Sigma^{-1} = \Phi\Lambda'\Sigma^{-1}.$$

This argument requires the identity $\Sigma - \Psi^2 = \Lambda\Phi\Lambda'$, which is immediate in FA; cf. (11). But at present, we can only derive: $\Sigma = \Lambda E[\hat{\mathbf{f}}\hat{\mathbf{f}}']\Lambda' + \Lambda E[\hat{\mathbf{f}}\mathbf{e}'] + E[\mathbf{e}\hat{\mathbf{f}}']\Lambda' + E[\mathbf{ee}']$. By the underdetermination theorem (15), it follows that in general, $E[\hat{\mathbf{f}}\hat{\mathbf{f}}'] \neq \Phi$; e.g., in the simple example used above, $E[\hat{\mathbf{f}}\hat{\mathbf{f}}'] = .74 \neq 1 = \Phi$. Also, for the reduction of $E[\mathbf{ee}']$ to Ψ^2 amounts to (FAiii). Similarly, if the middle two terms vanish because $\hat{\mathbf{f}}$ is the orthogonal projection of $\mathbf{f}$ onto $\mathbf{x}$, then the lemma appears to assume the very thing it purports to show. The argument also assumes $(\Lambda'\Lambda)^{-1}$ exists, which is not guaranteed.

[10] I leave aside the important but separate issue of missing data, which in the present case would supply only part of a new row or column.

[11] As usual, in some cases, various specific considerations can be of help. E.g., for f^5 not to differ in any important way from f, it is presumably necessary but not sufficient that they have sufficiently similar λs for the four xs that they share; Widaman [2007]. Here confirmatory factor analysis can help test whether there are statistically significant differences. More abstractly, there are some asymptotic results to the effect that roughly speaking, if you can add more variables that both correlate well with

f but are not too correlated with themselves or the existing xs, you can more tightly constrain the underdetermination—all the way to zero in the limit [Guttman, 1953, Mulaik and McDonald 1978].

[12] That is, the general situation can be characterized within an infinite dimensional vector space consisting of all random variables (Borel-measurable functions) with finite second moments—i.e., the only probabilistic properties that matters are those that they have *qua* abstract vectors in this space. The inner product in this space can be defined as the covariance between variables; cf [Guttman, 1955, Eaton, 1983].

[13] Norton is appropriately guarded in his language, and I don't mean to play gotcha with a counterexample, but only to investigate where FA stands with respect to this characterization of underdetermination.

[14] Such talk of degrees of underdetermination should not be confused with the very different logical taxonomy developed in Laudan [1990].

[15] At points, Guttman erroneously describes the variance of the *error* of a regression estimate as $R^2\sigma^2$; but this of course is the variance of the *estimate* itself, not the error, which has variance $(1 - R^2)\sigma^2$. However, the final result, given in Guttman's (44) of that page, is correct.

[16] The minimum correlation in (9) applies only to pairs of individual latent variables. More generally, Guttman's Theorem 3 [Guttman, 1955, 73] establishes the existence of a $\mathbf{f}^*$ (for a given $\mathbf{f}$) such that the norm of $\mathbf{f}^* - \mathbf{f}$ is maximized.

[17] If r is the minimum correlation between a factor f and some equivalent factor, then for all values $s \in [1, r]$, there exists another equivalent factor whose correlation with f is s. So these negative minimum correlations also imply the existence of factors that are uncorrelated with f.

[18] It might be possible to cook up particular examples where u is a nonlinear transformation of some of the xs, so that in certain specific cases, although u remains uncorrelated, it nevertheless aids in f's overall predictive capacity. While unusual, they don't affect the point made above. In actual practice, it would be most natural to include such a nonlinear transformation in the manifest data as another x (or to use something other than FA).

[19] Similarly, "If all we wanted was to maximize probability, we should never venture beyond our data" [Lipton, 2004, 110]. It is nearly impossible to find a sharper example of this than the contrasts between $\hat{f}$ and f.

[20] Sober writes: "the empiricist's preoccupation with sense experience takes the form of a thesis about the role of observation in science and the rationalist's emphasis on reason is transformed into a claim about the indispensable role of the *super-empirical virtues*" [Sober, 2008, 129; cf. 135]. Similarly, for Cartwright: "I have sometimes summarized my view about explanation this way: no inference to best explanation; only inference to most likely cause" [Cartwright, 1983, 6].

[21] (U1) most directly attaches to the estimation of the factor loadings. E.g., a typical run with $n = 10$, Λ for (3) was estimated as $[-.27, .49, -.40, .99]'$; such deviations from $[.8, .6, .5., .4]'$ effectively never occur when $n = 10,000$.

[22] This scenario can be contrasted with the case of measurement bias across groups; e.g. Millsap [2011].

[23] Likewise, "Although the factor analysis model is defined at the structural level, it is undefined at the data level. This is a well known but little discussed problem with factor analysis." [Revelle, 2013, documentation, p. 110].

[24] Since the distribution of $\hat{f}$ might be estimable from that of the xs, some economy can be had if f is first decomposed into its two parts, and Chebyshev's inequality is applied only to the unknown u.

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(5): 17–21.

Baird, D. (1987). Exploratory factor analysis, instruments and the logic of discovery. *British Journal for the Philosophy of Science*, 38(3): 319–337.

Bartholomew, D. J. (1981). Posterior analysis of the factor model. *British Journal of Mathematical and Statistical Psychology*, 34: 93–99.

Bartholomew, D. J. (1996). Comment on: Metaphor taken as math: Indeterminacy in the factor model. *Multivariate Behavioral Research*, 31(4):551–554, 631–636.

Bartholomew, D. J., Deary, I. J., and Lawn, M. (2009a). A new lease of life for Thomson's bonds model of intelligence. *Psychological Review*, 116(3): 567–579.

Bartholomew, D. J., Deary, I. J., and Lawn, M. (2009b). The origin of factor scores: Spearman, Thomson and Bartlett. *British Journal of Mathematical and Statistical Psychology*, 62: 569–582.

Bartholomew, D. J., Knott, M., and Moustaki, I. (2011). *Latent Variable Models and Factor Analysis: A Unified Approach*. John Wiley and Sons, Ltd., London, 3rd edition.

Basilevsky, A. (1994). *Statistical Factor Analysis and Related Methods*. Wiley-Interscience, New York.

Beauducel, A. (2013). Taking the error term of the factor model into account: The factor score predictor interval. *Applied Psychological Measurement*, 37(4): 289–303.

Bollen, K. (1989). *Structural Equations with Latent Variables*. John Wiley and Sons, New York.

Brereton, R. G. (2003). *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. John Wiley and Sons, Inc., Hoboken, NJ.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Clarendon, Oxford.

Eaton, M. L. (1983). *Multivariate Statistics: A Vector Space Approach*. Institute of Mathematical Statistics, Beachwood, OH.

Fabrigar, L., MacCullum, R., Wegener, D., and Strahan, E. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3): 272–299.

Fodor, J. (1974). Special sciences: Or the disunity of science as a working hypothesis. *Synthese*, 28: 97–115.

Fodor, J. A. (1975). *The Language of Thought*. HUP, Cambridge, MA.

Glymour, C. (1998). What went wrong? reflections on science by observation and *The Bell Curve*. *Philosophy of Science*, 65(1): 1–32.

Gould, S. J. (1996). *The Mismeasure of Man*. W.W. Norton and Co., New York, revised and expanded edition.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, 6(4): 430–450.

Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika*, 18(4): 277–296.

Guttman, L. (1955). The determinacy of factor score matrices with implications for five other basic problems in common-factor theory. *The British Journal of Statistical Psychology*, VIIII: 65–81.

Heermann, E. F. (1964). The geometry of factorial indeterminacy. *Psychometrika*, 29(4): 371–381.

Heermann, E. F. (1966). The algebra of factorial indeterminacy. *Psychometrika*, 31(4): 539–543.

Hunt, E. B. (2011). *Human Intelligence*. Cambridge University Press, Cambridge.

Jolliffe, I. (2010). *Principal Component Analysis*. Springer-Verlag, New York, second edition.

Kasper, D. and Ünlü, A. (2013). On the relevance of assumptions associated with classical factor analytic approaches. *Frontiers in Psychology*, 4: 1–20.

Kestelman, H. (1952). The fundamental equation of factor analysis. *British Journal of Psychology, Statistical Section*, 5: 1–6.

Kukla, A. (1996). Does every theory have empirically equivalent rivals? *Erkenntnis*, 44(2):137–166.

Kukla, A. and Walmsley, J. (2004). A theory's predictive success does not warrant belief in the unobservable entities it postulates. In Hitchcock, C., editor, *Contemporary Debates in Philosophy of Science*, pages 133–148. Blackwell, Oxford.

Laudan, L. (1990). Demystifying underdetermination. *Minnesota Studies in the Philosophy of Science*, 14: 267–297.

Ledermann, W. (1938). The orthogonal transformation of a factorial matrix into itself. *Psychometrika*, 3: 181–187.

Leplin, J. (2004). A theory's predictive success can warrant belief in the unobservable entities it postulates. In Hitchcock, C., editor, *Contemporary Debates in Philosophy of Science*, pages 117–132. Blackwell, Oxford.

Lipton, P. (2004). *Inference to the Best Explanation*. Taylor and Francis Group, London.

Lipton, P. (2008). Inference to the best explanation. In Psillos, S. and Curd, M., editors, *The Routledge Companion to Philosophy of Science*, pages 193–202. Routledge, London.

Lovie, P. and Lovie, A. (1995). The cold equations: Spearman and Wilson on factor indeterminacy. *British Journal of Mathematical and Statistical Psychology*, 48: 237–253.

Maraun, M. D. (1996). Metaphor taken as math: Indeterminacy in the factor analysis model (with discussion). *Multivariate Behavioral Research*, 31(4):517–538, 539–689.

McDonald, R. P. (1974). The measurement of factor indeterminacy. *Psychometrika*, 39(2): 203–222.

Millsap, R. E. (2011). *Statistical Approaches to Measurement Invariance*. Taylor and Francis, New York.

Mulaik, S. A. (2010). *Foundations of Factor Analysis*. Chapman and Hall/CRC, Boca Raton, FL, second edition.

Mulaik, S. A. and McDonald, R. P. (1978). The effect of additional variables on factor indeterminacy in models with a single common factor. *Psychometrika*, 43(2): 177–192.

Norton, J. (2008). Must evidence underdetermine theory? In Carrier, M., Howard, D., and Kourany, J., editors, *The Challenge of the Social and the Pressure of Practice*, pages 17–44. University of Pittsburgh Press, Pittsburgh, PA.

O'Hagan, A. and Forster, J. (2004). *Kendall's Advanced Theory of Statistics, volume 2B: Bayesian Inference*. Hodder Arnold, London, 2nd edition.

Pearl, J. (2000). *Causation: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge.

Piaggio, H. (1931). The general factor to spearman's theory of intelligence. *Nature*, 127(3193): 56–57.

Piaggio, H. T. H. (1933). Three sets of conditions necessary for the existence of a *g* that is real and unique except in sign. *British Journal of Psychology*, 24(1): 88–105.

Quine, W. v. O. (1951). Two dogmas of empiricism. In *From a Logical Point of View*, pages 20–46. Harvard University Press, Cambridge, MA, 2nd edition.

Quine, W. v. O. (1955). Posits and reality. In *The Ways of Paradox and Other Essays*, pages 233–241. Random House, New York.

Quine, W. v. O. (1972). Methodological reflections on current linguistic theory. In Davidson, D. and Harman, G., editors, *Semantics of Natural Language*, pages 442–54. D. Reidel, Dordrecht.

Quine, W. v. O. (1975). On empirically equivalent systems of the world. *Erkenntnis*, 9: 313–328.

R Core Team, (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reichenbach, H. (1956). *The Direction of Time*. UCLA Press, Los Angeles.

Revelle, W. (2013). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 1.3.2.

Schönemann, P. H. (1971). The minimum average correlation between equivalent sets of uncorrelated factors. *Psychometrika*, 36(1): 21–30.

Sober, E. (1984). Common cause explanation. *Philosophy of Science*, 51(2): 212–241.

Sober, E. (2008). Empiricism. In Psillos, S. and Curd, M., editors, *The Routledge Companion to Philosophy of Science*, pages 129–138. Routledge, London.

Spearman, C. (1904). 'general intelligence', objectively determined and measured. *The American Journal of Psychology*, 15(2): 201–292.

Spearman, C. (1922). Correlation between arrays in a table of correlations. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 101(708): 94–100.

Spearman, C. (1927). *The Abilities of Man: Their Nature and Measurement*. Macmillan and Co., London.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*. MIT Press, Cambridge, MA, 2nd edition.

Steiger, J. H. and Schönemann, P. H. (1978). A history of factor indeterminacy. In Shye, S., editor, *Theory Construction And Data Analysis in the Behavioral Sciences*, pages 136–178. Jossey Bass, San Francisco.

Sternberg, R. J. and Kaufman, S. B., editors, (2011). *The Cambridge Handbook of Intelligence*. Cambridge University Press, Cambridge, UK.

Suppes, P. and Zanotti, M. (1981). When are probabilistic explanations possible? *Synthese*, 48: 191–199.

ten Berge, J. M. F., Krijnen, W. P., Wansbeek, T., and Shapiro, A. (1999). Some new results on correlation-preserving factor scores prediction methods. *Linear Algebra and its Applications*, 289: 311–318.

Thomson, G. H. (1916). A hierarchy without a general factor. *British Journal of Psychology*, 8(3): 271–281.

Thurstone, L. (1934). The vectors of mind. *Psychological Review*, 41: 1–32.

Thurstone, L. (1938). *Primary Mental Abilities*. University of Chicago Press, Chicago.

Thurstone, L. (1947). *Multiple Factor Analysis*. University of Chicago Press, Chicago.

Uffink, J. (1999). The principle of the common cause faces the Bernstein paradox. *Philosophy of Science*, 66: S512–S525.

van Fraasen, B. C. (1980). *The Scientific Image*. Oxford University Press, Oxford.

Widaman, K. (2007). Common factors versus components: Principals and principles, errors and misconceptions. In Cudek, R. and MacCallum, R. C., editors, *Factor Analysis at 100*, pages 177–203. Lawrence Erlbaum Associates, Mahwah, NJ.

Wilson, E. B. (1928a). On hierarchical correlation systems. *Proceedings of the National Academy of Sciences*, 14(3): 283–291.

Wilson, E. B. (1928b). Review of *The Abilities of Man, Their Nature and Measurement* by C. Spearman. *Science*, 67(1731): 244–248.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, Oxford.