# What psychologists should know about the philosophy of science

Review of *Investigating the Psychological World*:

*Scientific Method in the Behavioral Sciences* (2014)

Brian D. Haig, Cambridge, MA: Bradford/MIT.

Gregory J. Feist

Department of Psychology

San Jose State University

and

Michael E. Gorman

School of Engineering & Society

School of Engineering and Applied Science

University of Virginia

Brian Haig's *Investigating the Psychological World*: *Scientific Method in the Behavioral Sciences* puts forth a new theory of method for the behavioral sciences, namely an abductive theory of method (ATOM). Method is how scientific knowledge comes to be and yet knowledge and interest in method are relatively lacking in current graduate school training. Haig's central thesis is that progress in the behavioral sciences is being hindered by its lack of understanding a theory of method that goes beyond the classic inductive or hypothetico-deductive theories of the past and on which much of current science is based. To be sure, the ATOM model is deeply grounded in modern philosophy of science and the book is much a work in the philosophy of science as it is in research methodology (see also, Haig, 2005).

The core tenet of Haig's theory of method is that scientific knowledge flows from detecting phenomena to generating, developing, and appraising theories. The addition of the corresponding strategies, methods, and inferences to each stage of science building makes the model all the more compelling. The ATOM model is built around non-experimental techniques such as partial correlation, multiple regression, exploratory and confirmatory factor analysis, and meta-analysis. Therefore, it most easily and readily applies to non-experimental methods in the behavioral sciences. It is not clear, except in very abstract and implicit form, how this model may be applied to experimental techniques of manipulated variables and random assignment to experimental and control conditions.

Science is guided by problems whose solutions are constrained by methods, theory, and observation. Scientists analyze results searching for more than findings: they search for phenomena (robust empirical regularities). Once these phenomena are uncovered they are explained abductively by inferring their causes. The major goal of ATOM in particular

and science in general is twofold: 1) to detect robust phenomena; 2) and to explain these phenomena through abductive theory construction.

To overview ATOM, there are three Phases (strategies, methods, and inferences) and two periods: Phenomenon Detection and Theory Construction. Each Phase is broken into its two periods, making a 3 x 2 model. For example, phenomenon detecting strategies include controlling for confounds and replicating findings; methods include preliminary and exploratory data analysis, simulated resampling, and meta-analyses; and inferences include enumerative induction, which involves empirically derived knowledge of when and how generalized phenomena can be applied on a case-by-case basis.

Theory Construction in turn has three subperiods: generation, development, and appraisal, each with their corresponding strategies, methods, and inferences. For example, a strategy for constructing a theory may involve creating a rudimentary and plausible theory, which is developed through analogical modeling. A strategy for evaluating that theory in turn would involve a direct comparison to a rival theory. A method for generating a theory is exploratory factor analysis, and a method for developing it might involve evaluating analogies from source models. Theory appraisal methods may involve structural equations modeling or an evaluation of how coherent the explanation is.

Detecting Phenomena

The second chapter focuses on phenomenon detection, with a special eye toward psychological phenomena. As mentioned above, *phenomena* are stable, recurrent and general features of the world. They are latent more than manifest, abstract more than concrete and they are not directly observable. As examples of phenomena in psychology, Haig offers the Law of Effect, the Flynn Effect, and the Recency Effect.

To be sure, phenomena ultimately stem from data but are not data themselves. *Data* are less stable, idiosyncratic and context/study specific. Data are not yet established into laws, principles, or effects. In order for a set of data to become a phenomenon certain conditions must be met. In general terms, only high quality, robust data get promoted to phenomena. More specifically, the promotion happens when results from individual studies make heavy use of controls for extraneous variables (confounds), when the results have been replicated many times, aggregated across numerous studies (as in meta-analyses), when results have been triangulated, and calibration of measures have been carried out. These last two points require some elaboration. Triangulation occurs when results converge on a similar outcome using many different techniques, samples, and even measures. In psychometrics, this is known as convergent validity. Calibration occurs when a measure used in a study is first calibrated against a well-established yardstick. Calibration is common in the physical and even biological sciences when machines are involved, but Haig does not really describe in adequate detail how social science measures (i.e., self-report measures) might be calibrated. Obviously, any good self-report measure must be validated against established metrics and assess the construct in question. This is what construct validity is about.

Another issue Haig addresses in this context is his distinction between pseudo and real phenomena. Like pseudo-science, pseudo-phenomena on first glance may appear to be real, but in fact do not meet the criteria of being robust and stable. In the history of psychology, a few examples of pseudo-phenomena would be Watson's Little Albert (conditioned reflex) and Herrnstein's claim that IQ "is psychology's best proved, socially significant empirical finding" (p. 53). Haig's claim that the heritability of intelligence is a

pseudo-phenomenon is based on the controversial findings of one of its first proponents, Cyril Burt. This is one place where Haig is factually wrong. Burt's data may well have been bogus (although even that claim is somewhat in dispute), but that is irrelevant to the dozens of more modern studies on the heritability of IQ that demonstrate beyond a doubt that IQ is heritable and the general conclusion is that about 60% of the variation in intelligence is due to genetic influence. Indeed, these studies have been well-replicated, have controlled for confounds, and have been aggregated across many studies using different measures (e.g., Bartels, Rietveld, van Baal, & Boomsma, 2002; Bouchard & McGue, 1981; Grigorenko, 2000; Lynn, 2006; Plomin & Petrill, 1997). Such a misleading statement is all the more surprising coming from a writer who is clearly well versed in the history of intelligence and argues that another somewhat less robust finding in the intelligence literature—the Flynn Effect—is a prototypic example of a phenomenon.

Theory Construction: Generation

In the third chapter, Haig discusses the role of exploratory factor analysis (EFA) in theory generation and argues EFA and ATOM in general offer a better alternative for generating scientific theories than the more commonly accepted inductive and hypothetico-deductive methods. Haig attempts two major goals in this chapter: first, he describes how EFA and the principle of common cause are abductive methods for generating scientific theories; second, he argues that ATOM along with both EFA and confirmatory factory analysis (CFA) can and should be used together to solve problems such as indeterminacy (i.e., when factors or theories are not adequately determined by empirical evidence).

Among the general points Haig makes in the first section of the chapter are that abductive inference is used both to detect phenomena and to justify attempts to further

test their plausibility. In addition, multivariate statistical procedures (such as partial r and multiple regression) are used in abductively going from correlational patterns to plausible initial theory using common cause as a guide. The main goal of EFA is to generate explanatory theories that explain correlational patterns (abductively).

Haig begins the chapter by explicating the nature of modern scientific abductive inference and even gives its general form [and uses Spearman's *g* factor model as a specific example]:

> "The surprising empirical phenomenon, P [positive manifold], is detected.
>
> "But if Hypothesis H [*g*] were approximately true, and the relevant auxiliary knowledge, A, were invoked [and measured reliably and validly by WAIS], then P [positive manifold] would follow as a matter of course.
>
> Hence there are grounds for judging H [g] to be initially plausible and worthy of further pursuit." [p. 63].

Along with EFA, the principle of common cause is an important means of abductively generating plausible initial theories or hypotheses. Common cause is a form of abductive reasoning because a person reasons from correlated events to common causes believed to explain those correlations. "Whenever two events are improbably, or significantly, correlated, we should infer a common cause, unless we have good reason not to." (p. 68)

One criticism of EFA-based theory, however, is it does not explain phenomena. At this point, as a personality-trained psychologist, Greg Feist would have enjoyed seeing Haig instantiate his argument with the history of the Big Five Model of personality and its development from of a linguistic analysis of trait terms (Allport, 1937), to factor-analytically derived scales (Cattell, 1957; Costa and McCrae, 1985; Eysenck, 1952; John, 1990) to a theory (McCrae and Costa, 2008). Indeed, for years and years, the main short-coming of this factor-analytically derived model of personality was that it was purely descriptive and not at all explanatory. It has only been in the last 10 years or so that McCrae and Costa have developed an explanatory theory of the structure of personality (McCrae & Costa, 2008).

Factor indeterminacy occurs when observed variables belong to more than one latent factor, that is, when different factor structures fit the same correlation pattern found in observed/manifest variables. Indeterminacy prevents us from determining the truth or falsity of a hypothesis or theory based on available evidence. In short, we are unable to choose one theory from among many. Haig refers to this as "underdetermination of theory by empirical evidence" (UTEE). Indeterminacy exists and yet can be dealt with by realizing that EFA can narrow down the plausible suspects. In fact, the principle of strong inference or as Haig describes it theoretical pluralism is crucial here: "Thus it is through the simultaneous pursuit of multiple theories with the intent of eventually adjudicating between a reduced subset of them that one arrives at judgments of best theory." (p. 79).

Theory Construction: Development

Analogical modeling is critical to the author's ATOM approach. He develops this topic by first considering modeling. Duhem and Feigl were skeptical of models, and so historically

were psychologists, but models now are considered a major scientific tool. The author's examples mostly come from cognitive science, which opened up the Skinnerian black box to look at what processing must be going on. Climate science, biomedical engineering and other fields that deal with complex systems are particularly model-intensive.

The author discusses several types of models but focuses primarily on analogies. He uses Darwin's deep understanding of breeding as an example: if new traits can be developed through breeding, then perhaps similar processes account for new species in nature. But the author points out this is not how Darwin actually worked; Darwin used natural selection to refine his understanding of breeding. The author also presents Harre's analogy between theater and human interaction: perhaps human behavior is like being on a stage. But just like the Darwin case, the Harre example works both ways: theater is modeled on human social interaction. The book would be enhanced by more examples of scientific reasoning like Darwin's and Harre's because these cases show the inadequacies of the philosophical models.

Theory Construction: Appraisal

According to Haig, Inference to the Best Explanation (IBE ) follows this general form:

D is a collection of data.

Hypothesis H explains the pattern in the data.

No other hypothesis explains D as well as H

Therefore H is likely to be true.

Abduction occurs when D involves a surprising phenomenon.

Haig mentions the idea that the best explanation is the loveliest. As Paul Dirac said, "It's most important to have a beautiful theory. And if the observations don't support it,

don't be too distressed, but wait a bit and see if some error in the observations doesn't show up" (see Gorman, 1992).

Semmelweis' discovery of the need for sterile procedures is described as an example of contrastive explanation: Semmelweis compared a ward where he introduced hand-washing to one where it was not done, and noted the drop in childbed fevers.

Thagard's Explanatory Coherence is one method for achieving IBE. Explanatory coherence is determined by:

- explanatory breadth, or accounting for a greater range of facts than other competing theories
- simplicity, or introducing fewer ad hoc assumptions
- analogies to other theories scientists find credible

Thagard provided proof-of-concept for EC by creating a computational simulation (ECHO) that could be used to apply IBE to historical competitions between theories like the superiority of Darwin's theory to creationist alternatives.

Haig includes a nice section on the relationship between IBE and structural equation modeling, than turns to criticisms. Van Fraasen asked whether IBE merely selects the theory that is the 'best of a bad lot'. ATOM solves this problem by judging theories based on their initial plausibility before comparing them using IBE.

Bayesianism is usually viewed as an alternative to IBE because of the former's focus on prior probabilities in theory evaluation. IBE would certainly work better than Bayes in situations where prior probabilities were not available. For psychological researchers, the bottom line is to evaluate their theories based not only on fit with their data but also on other criteria provided by IBE and explanatory coherence.

Conclusion

The final chapter includes a laundry-list of issues related to ATOM, some new, most review. ATOM incorporates a constraint-inclusion approach to identifying problems, in which the constraints are constitutive of the problem in some way that is not explained. Constraints can be metaphysical, methodological and/or based on relevant scientific knowledge. ATOM is both generative and consequentialist in terms of method, which means ATOM considers both the means by which evidence is obtained and the strategies used to test the results against other hypotheses. For justification, ATOM would recommend methods like replication and resampling to determine how reliable a result was, and would apply approaches like Thagards' explanatory coherence to determine if an explanation for a result fits in with other reliable results. There is no mention of a paradigm shift, where a connected network of theories and results can be transformed into a new pattern.
The author is best when he speaks directly to behavioral scientists, as when he argues that behavioral science relies too much on a hypothetico-deductive approach and thinks too little about theory construction. Here again, examples would be most helpful. Consider cognitive dissonance, a very simple theory that led to a huge program of inconclusive hypothetic-deductive research. Had more attention been paid to constructing a better theory initially, and continuing to refine it, less time would have been wasted on trying to test it.

Haig sees grounded theory as a process of abducting theories from data, elaborating the theory by developing models and justifying the emerging theory by its explanatory coherence. This process should be iterative, in the sense that theory and models abducted

from data should be applied to more data and continually revised. Researchers should be on the look out for better perspectives as they go through this process.

According to Haig, ATOM can be a part of the kind of reasoning done by a clinical psychologist coming up with a diagnosis. When making a diagnosis, a clinician collects data, usually through an interview, then goes through ATOM's four stages before writing up the case. Haig fails to emphasize is that clinical judgment evolves constantly over multiple encounters with a client, and that treatment itself is a series of experiments to see what is effective. Diagnosis is important to make certain the client's insurance will pay--so diagnosis is a political and social act. Haig includes a section attempting to rebut two earlier critiques of this approach, mainly on the grounds that theories cannot be determined by evidence alone. This section might have been better in an appendix.

The book ends by briefly considering lessons for the education of psychologists, who would benefit by taking a more critical attitude towards the methods they use. We would actually use the term reflective instead of critical. The ability to reflect on prevailing methodological catechisms is liberating: it leads to the development of new methods like grounded theory that address limitations of other approaches.

To be sure, Haig's writing is rather jargon-filled and very deeply grounded in philosophy so it is not for the philosophical novice. Therefore, reading the book is not an easy or light undertaking. Moreover, many of the same ideas and arguments are made multiple times so in fact the book could be sharpened and streamlined a bit more.

Writing style critiques aside, the book is a clear and important contribution to the science of science in general and to the theory of method in particular. Haig is correct to argue the behavioral sciences would progress in a more systematic and deliberate fashion

were the majority of its practitioners to adopt a model such as ATOM. For that reason, we recommend this book to any graduate students, post-docs, and academic scientists in psychology and in most of the behavioral and social sciences interested in conducting research fully informed by the philosophical and methodological issues they are confronting. If implemented the abductive theory of method would advance the methodological foundation of research in the behavioral and social sciences.

References

Allport, G. W. (1937). *Personality: A psychological interpretation.* New York: Henry Holt.

Bartels, M., Rietveld, M. J., van Baal, G. C., & Boomsma, D. I. (2002). Genetic and environmental influences on the development of intelligence. *Behavior Genetics, 32,* 237–249.

Bouchard, T. J., Jr., & McGue, M. (1981). Familial studies of intelligence: A review. *Science, 212*, 1055–1059.

Cattell, R. B. (1957). *Personality and motivation structure and measurement.* New York: World Book.

Costa, P.T., & McCrae, R.R. (1985). *The manual for the NEO Personality Inventory*. Odessa, FL: Psychological Assessment Resources.

Eysenck, H.J. (1952). *The structure of human personality.* London: Methuen.

Gorman, M. E. (1992). *Simulating science: Heuristics and mental models in technoscientific thinking*. Bloomington: Indiana University Press.

Grigorenko, E. (2000). Heritability and intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 53–91). New York, NY: Cambridge University Press.

Haig, B. D. (2005). An abductive theory of method. *Psychological Methods, 10,* 371-388.

John, O.P. (1990). The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L.A. Pervin (ed.), *Handbook of personality: Theory and research* (pp. 66-100). New York: Guilford Press.

Lynn, R. (2006). *Race differences in intelligence: An evolutionary analysis.* Augusta, GA: National Summit.

McCrae, R.R., & Costa, P. T. (2008). The Five-Factor theory of personality. In O.P. John, R.W.

Robins, & L. A. Pervin (Eds.) *Handbook of Personality: Theory and research* (3rd edition). (pp. 159-181). New York: Guilford Press.

Plomin, R., & Petrill, S. A. (1997). Genetics and intelligence: What's new? *Intelligence, 24,* 53–77.