

Running head: ATOM's normative status

The All-Too-Flexible Abductive Method:

ATOM's Normative Status

Jan-Willem Romeijn

University of Groningen

Abstract

This paper discusses the abductive theory of method (ATOM) by Brian Haig from a philosophical perspective, connecting his theory with a number of issues and trends in contemporary philosophy of science. It is argued that as it stands, the methodology presented by Haig is too permissive. Both the use of analogical reasoning and the application of exploratory factor analysis leave us with too many candidate theories to choose from, and explanatory coherence cannot be expected to save the day. I end with some suggestions to remedy the permissiveness and lack of normative force in ATOM, deriving from the experimental practice within which psychological data are produced.

Keywords: scientific methodology, abduction, theory generation, theory development, underdetermination, exploratory factor analysis, analogical reasoning, Newman's problem, experimental interventions.

Setting the Stage: Normative Methodology

Ever since the advent of modern science, or natural philosophy as it was first termed, those left behind in traditional philosophy have investigated the status of scientific knowledge. The emergence of this new type of knowledge posed new philosophical problems, since the knowledge was not just rooted in explicit metaphysical and epistemological positions but, at least in part, in a direct contact with the empirical world. On the one hand, attempts were made to put science on a proper footing, either by rationalist or by empiricist principles. But the roots of science in empirical reality were also severely criticized, most famously in the problem of induction by David Hume. He argued that the presence of regularities in the empirical world is never a solid basis for supposing some structure behind the regularity, or even for expecting the regularity to continue in the future.

I think it is fair to say that Hume's arguments set the stage for the separate philosophical discipline called 'philosophy of science', associated mostly with the logical positivists such as Neurath, Schlick and Carnap. The logical positivist project was to provide a basis, and ultimately a justification, of scientific fact, thereby resolving the Humean criticisms. Logical positivists organized the investigation of scientific fact roughly around two questions, one dealing with the demarcation of science against pseudo-scientific activities such as astrology and augury, the other dealing with the justification of science thus defined. The leading idea was to answer both these questions by an analysis of scientific method, and thus to answer both of them in one fell swoop: facts arrived at by the appropriate method were deemed scientific, and because they were

arrived at by this method, scientific facts were also justified. So both in demarcation and in justification, the study of scientific method took centre stage.

The twentieth century saw the emergence of a large number of scientific methodologies: the inductivist views of Carnap, the falsificationist theory of Popper, the hypothetico-deductive method of Hempel, the naturalised methodology of Laudan, the methods of truth approximation by Niiniluoto and Kuipers, the AI-inspired methodology by Thagard, the Bayesian approach by Howson and Urbach, the defences of inference to the best explanation by Lipton and Psillos, and the more recent confirmation theory of formal epistemologists can all be viewed as attempts to provide a philosophical basis and justification of scientific fact. It is notable that all these attempts have a normative nature: in some way or other they determine a proper scientific method and thus provide a recipe for arriving at justified scientific facts. It must also be noted that philosophers have not always had an open eye for methodological developments in the sciences themselves, while working scientists have not always been sensitive to the subtlety, or perhaps pettiness, of philosophical argumentation. With the exception of Popper, who has had a marked influence on the behavioral and social sciences, scientists and philosophers of science have by and large kept to their own.

Within this philosophical and scientific setting, the abductive theory of method (ATOM) presented by Brian Haig (2005a) is a timely and very welcome contribution. One of the main merits of this new methodology, I feel, is that it has an open eye for the practice of science. It is driven not so much by methodological considerations internal to the philosophy of science, but rather by the methods that working psychologists rely on. Meanwhile, it is a methodological theory in line with other such theories, and so it is a

contribution to the philosophical discussion of scientific method. As such, Haig's ATOM may be a means for bringing scientists and philosophers closer together in their aims to characterize and justify science.

Yet the descriptive merits of ATOM are also the source of a potential problem, which I will spell out in more detail in this article. As indicated in the foregoing, the philosophical reflection on science, especially in its focus on methodology, has been geared towards both a characterization and a justification of scientific facts. One of the defining worries in the philosophy of science is the problem of induction, and a methodology may be expected to provide an account of how this problem can be resolved. So if we take ATOM as a methodology, we may require not just that it adequately characterizes science, but also that it provides a normative account of science in which the problem of induction is either resolved, or otherwise elegantly by-passed. The central argument of this article is that as it stands, ATOM cannot live up to this challenge.

This is not to say that ATOM is not in a position to adapt in order to meet the challenge. In fact, I will argue that ATOM is in exactly the right position to do so, just because it has an open eye for scientific practice. Most methodologies in the philosophy of science restrict attention to the context of justification: the theories have already been formulated, and all the data has been gathered and put in the appropriate format to function as evidence, or to be called a 'phenomenon', after which the justified scientific method can kick in. So methodological norms are only applied after theory and phenomenon have been put in place. But for ATOM, both the transformation from raw data to phenomenon and the generation of the theory which is to be submitted to testing

are taken to be part of the story. That is, ATOM is concerned with both the context of justification and the context of discovery. I will argue that a deeper analysis of the context of discovery may provide ATOM with the means to face the problem of induction.

Before spelling out the problem of induction in more detail, let me summarize the two central claims of this paper. I indicated that where more traditional methodologies restrict attention to the context of justification and provide normative guidelines within this restricted context, ATOM also addresses methodological issues in the context of discovery. Now my first claim is that to its disadvantage, ATOM cannot adequately fulfill a normative role, and my second claim is that extending ATOM in the direction of the context of discovery will remedy this shortcoming. Even stronger, I want to suggest that with its interest in the context of discovery ATOM finds itself at the origin of an emerging philosophy of science that combines an open eye for scientific practice with a normative interest in, and an analytic perspective on, methodology.

The article is set up as follows. In section 2 I discuss the problem of induction in some more detail. The goal with this is to make precise what problems a scientific methodology, and ATOM in particular, is supposed to solve. In sections 3 and 4, I deal with two pieces of methodological advice from ATOM, to do with factor analysis and the use of analogy. In these sections I argue that as a normative theory, ATOM is too permissive. In section 5, finally, I consider explanatory coherence as a solution to this permissiveness, but also sketch the outline of an alternative solution drawing on an analytic characterization of scientific practice. My suggestions will make clear that I think ATOM is on the right track.

What's Next: The Problem of Induction

In this section I discuss the problem of induction, first as a problem with predictions on the basis of empirical facts, and then as a more general problem of the underdetermination of theory by empirical evidence. I further discuss statistical underdetermination as a special case of underdetermination generally. Finally, following van Fraassen (1989) and Psillos (1999), I briefly review the role that abductive inference has in relation to underdetermination.

Induction is a mode of inference that runs from empirical facts that have already been obtained to as yet unknown empirical facts. For example, we may have received the data that all subjects with a positive score on some test show some pattern of behavior. By induction we may then conclude, or predict, the datum that the next subject who scores positive will also show this behavior. It will be clear that both everyday and scientific knowledge are very often based on such inductive inferences: we administer drugs because we expect them to be effective on the basis of past performance, and we eat bread because it has been wholesome all our past lives. A justification of scientific facts will therefore have to include a justification of induction. However, close scrutiny of induction reveals that this mode of inference is very hard to justify.

At first it may seem that induction can be justified by the overwhelming success of inductive inferences. But this is circular: to say that this past success has any bearing future success is again an instance of inductive inference. In response to this we may concede that induction cannot be defended just like that, but that we may assume that certain properties of the world remain the same. However, this is only moving the

problem to another location, because now the problem is that we do not know which properties can be deemed invariant. And using induction to select those properties again comes down to assuming what we set out to defend. Now the problem of induction is exactly this embarrassing difficulty of justifying inductive inference. And since much of scientific fact is obtained by induction, the problem of induction is a problem for the justification of scientific facts as well.

The above way of presenting the problem of induction reveals its direct relation to another well-known problem with justifying scientific knowledge by empirical facts, namely the problem of underdetermination of theory. In order to generate predictions and explanations, a scientific theory will have to pin down which aspects of its subject matter may be considered invariant, necessary, or even lawlike, and which aspects are variable and contingent. Typically this choice is informed by empirical evidence: we choose laws and stable regularities on the basis of the evidence that the empirical facts present us. Now the problem of underdetermination is that, relative to the way we construct the empirical evidence, there may be indefinitely many scientific theories that are equally well supported. So if we decide between theories exclusively on the basis of empirical evidence, the decision is underdetermined. Another way to express this underdetermination is that there are indefinitely many sets of properties that we may deem invariant on the basis of the empirical facts. And since only those invariant properties may be used in induction, this comes down to saying that the empirical facts do not determine what inductive inference to make. Viewed from this angle, the problem of underdetermination is a particular version of the problem of induction.

Very well, but how is this problem applicable to psychological theory? Scientific theories can take various guises. They may consist of a collection of models and experimental practices, or of a set of sentences in a formal language, or of a specification of laws and mechanisms. The class of statistical theories consists of general statements in conjunction with statistical models. Due to the noisy nature of human behavior, psychological theories are often of this type. Within this class of statistical theories, the problem of underdetermination takes a specific form, which I think is particularly relevant in the area of psychology.

In fact there are at least two types of underdetermination at issue in statistical theories: one relating to unidentified models, and one relating to model selection. As for the first, when a theory has fixed its statistical model, the empirical data may be used to choose among the statistical hypotheses included in the model, for example by a maximum likelihood estimation, or by a generalized least squares estimation. The estimation criterion determines how we construct the empirical evidence from the data and, relative to this criterion, we may land ourselves in a problem of underdetermination because it may be that for any set of empirical data that could have been obtained, or for the specific data set obtained, there are multiple hypotheses in the model that perform best on the criterion. In the literature this is sometimes called the problem of equivalent models, but we must be careful with the terminology here. In the present paper a model is a collection of statistical hypotheses, and each hypothesis is associated with a probability function over the data. The underdetermination problem of unidentified models is therefore, in the terminology of the present paper, more aptly described as the problem of equivalent hypotheses.

A slightly different type of underdetermination occurs in model selection, namely when the theory is such that a number of different statistical models is consistent with it. For example, it may be that the theory allows for two statistical models, one with two and another with three parameters. In such a case the estimation procedure may tell us what the best hypothesis is within each of the two models, but it does not tell us which of the two best hypotheses to choose. Both in the situations of unidentified models and in the case of model selection, we can say that the theory is statistically underdetermined, meaning that the data and the statistical model do not single out a unique hypothesis, and hence do not specify a unique theory.

The problem of underdetermination is strongly related to the topic of this paper, abductive inference. As indicated in Haig (2005a), abduction is principally a mode of inference that takes us from empirical facts to a plausible hypothesis that explains these facts. At this point Haig refers to the ideas of Peirce, who characterized abductive inference roughly as follows: “A surprising phenomenon is observed. If the hypothesis were true, the phenomenon would be a matter of course. Hence the hypothesis is worthy of further pursuit.” But we can unpack this characterization a bit more to arrive at an important aspect of abductive inference, since the empirical facts may be “a matter of course” in various ways. They may be very likely, or even deductively entailed, in which case the empirical evidence points to the hypothesis directly. But the empirical facts may also be a “matter of course” because of additional explanatory virtues that the hypothesis has. That is, the hypothesis may be inferred abductively because it is the best explanation of the empirical facts. One of the aspects of abductive inference in science is that it

allows us to choose between hypotheses, or theories, on the basis of explanatory considerations, over and above the direct evidence that the empirical facts provide.

By virtue of this latter aspect, abductive inference can be used to adjudicate in cases of underdetermination. Abduction, as discussed in Lipton (2004), van Fraassen (1989), and Psillos (1999), is the mode of inference that enables us to make a justified choice between empirically equivalent theories. The idea is that a justification of scientific facts ultimately requires a justification of abductive inference. If the empirical evidence, as constructed from the empirical facts, does not suffice to make a choice among several rival theories, then we may use an abductive inference to choose the best explanation among the empirically equivalent rivals. The notion of best explanation hinges on theoretical and pragmatic criteria, such as simplicity, coherence with background theories, practical applicability, and so on. In the specific case of statistical underdetermination for model selection, for example, the use of the theoretical criterion of simplicity is very prominent. Insofar as model selection employs simplicity, we may therefore call it abductive inference.

With the problem of induction and its relation to abduction in mind, let us have a fresh look at ATOM. If it is to count as a methodology, in line with other methodologies in the philosophy of science, we may expect it to provide an answer to the problems of induction and underdetermination. The fact that ATOM takes abductive inference as its centre piece seems very promising. However, as I will argue in the following, ATOM fails to meet the challenge that is posed by underdetermination. To some extent it provides a characterization of the theoretical criteria that can serve in the abductions, but

the specific abductive inferences which it claims to be warranted, do not single out a unique theory.

Fear and Loathing in Factor Analysis: Exploratory Factor Analysis and Theory Generation

ATOM sets forth exploratory factor analysis as one of its submethods, in part to detect phenomena but mostly for generating theory. The following concerns exploratory factor analysis in this latter role. I characterize various types of underdetermination at stake in factor analysis, all of them versions of statistical underdetermination, and I argue that this underdetermination presents a serious problem to exploratory factor analysis when used for abductive inference.

Before getting to factor analysis in ATOM and the problem of underdetermination, let me emphasize that Haig is certainly aware of the problems sketched below. Haig's characterization of theory generation is quite open-ended in this respect. Both in his paper on exploratory factor analysis (2005b), and in his exposition of ATOM (Haig, 2005a), he points out that the other submethods of ATOM will go some way towards resolving the underdetermination issues. However, as I argue in sections 4 and 5, I do not find these resolutions entirely convincing, and so in section 5 I propose an alternative submethod for dealing with underdetermination, which I consider a rather natural extension of ATOM. In any case, the present section must not be read as a stand-alone criticism of Haig's methodology, but rather as part of the motivation of an alternative and supplemented submethod, to be presented later on.

Exploratory factor analysis is a technique that posits a specific statistical model of latent random variables on the basis of an analysis of the correlational structure of observed random variables; see Mulaik (1985) for a philosophically-minded introduction. Say that in some experimental setting we observe degrees of fear and loathing in a number of individuals, and we find a positive correlation between these two variables. One way of accounting for the correlation is by positing a statistical model over the variables in which fear and loathing are directly related, and then estimate the parameters in the model. But we may feel that this model does not capture the causal or mechanistic details of the experimental setup. It may be that it is not the loathing that instills fear in people, or the fear that invites loathing, but rather that both these feelings are caused by a drug that is administered in the experiment. The correct statistical model, we may argue, posits a correlation between the drug and the fear, and similarly a correlation between the drug and the loathing, while conditional on a certain drug dosage, fear and loathing are uncorrelated. We then say that the drug dosage is the common factor to the observed variables of fear and loathing. The correlations between drug dosage and fear and loathing respectively we call the factor loadings.

In the above experimental setting we can also observe the common factor of drug dosage directly. But in situations in which the causal or mechanistic story is unknown, we may nevertheless want to posit such an underlying story. For example, recurring feelings of fear and loathing may be two of a large number of variables on negative feelings, used to describe individuals in a general population. Now if all these variables are strongly positively correlated, it may be that we can account for all the correlations by a statistical model positing a fairly small number of common factors, or even a single common factor

such as depression. Exploratory factor analysis is a technique for doing the latter in a systematic way. When given a set of correlations among observed variables, it produces a statistical model of latent common factors that can account for these correlations and, which given specific values of the latent common factors, leaves the observed variables uncorrelated.

So far so good; it seems that exploratory factor analysis is indeed a natural tool for abductive inference. With a little imagination it generates a model of underlying causes or mechanisms on the basis of observations. But what has happened to the supposedly insurmountable problem of underdetermination? In the next few paragraphs I will argue that this problem has certainly not vanished. For one thing, there will generally be a large number of latent common factor models of varying complexity which will fit the data to differing degrees, and so there will have to be a trade-off between goodness of fit and model simplicity. In other words, exploratory factor analysis suffers from the underdetermination associated with model selection. But this need not surprise us too much. All statistical modeling must at some point address this worry. However, it turns out that even within each common factor model there is an underdetermination problem. This is the underdetermination associated with unidentifiable models discussed in the foregoing, but within factor analysis it takes various guises.

Naturally, exploratory factor analysis only makes sense if the number of latent variables is smaller than the total number of observed variables, but there is no need to restrict models to a single common factor. And given a factor model with multiple latent variables, we find that the matrix with factor loadings is not completely determined. If we allow the latent factors to be correlated, we may rotate the matrix with estimated factor

loadings to obtain a different solution, and each of these estimated factor loadings will perform equally well on the goodness-of-fit criterion, be it maximum likelihood or generalized least squares. That is, the fit criterion does not lead to a single best hypothesis in the factor model, but to a collection of them, and so factor analysis suffers from a problem of underdetermination due to unidentified models. One standard reaction to this problem is to adopt the theoretical criterion that the latent variables must be independent. In a philosophical assessment of factor analysis as a submethod of a scientific methodology, however, I want to bring to the fore the idea that some such theoretical criterion is needed. This criterion does not drop out of the data themselves.

Quite apart from the foregoing, there is yet another underdetermination problem with factor analysis; see McDonald (1974), Steiger (1979) and Maraun (1996) for discussions. Say that we have rotated the matrix of factor loadings to meet the theoretical criterion of our choice, for instance by assuming a single common factor or by fixing the independence of the latent factors. Can we then reconstruct the latent variable itself, that is, can we provide a metric ordering so that each individual is assigned a determinate expected latent score? Sadly, the answer here is negative. We still have to deal with the so-called indeterminacy of the factor scores, meaning that there is a variety of ways in which we can organize the allocation of the individuals on the latent scores, all of them perfectly consistent with the parameter estimations. The type of underdetermination that is presented by this specific indeterminacy depends on what we take the exact statistical model underlying factor analysis to be. If we take the factor analysis model to specify a complete probability assignment over the latent and observed variables, then factor indeterminacy points to an underdetermination of these full probability assignments. In

that case we can say that factor analysis suffers from statistical underdetermination in the sense of unidentified models in two ways, one concerning the rotation freedom in the matrix of factor loadings, and one concerning factor score indeterminacy.

In sum, there are several problems of statistical underdetermination with factor analysis. We can now reconsider exploratory factor analysis as a submethod of the abductive theory of method in light of this underdetermination. The first thing that comes to mind is that this underdetermination of exploratory factor analysis in ATOM sits badly with one of the main functions of abductive inference in science. Recall that abduction is traditionally used to choose between rival theories if the empirical facts do not determine a unique best theory. That is, abduction is a mode of inference that allows us to employ theoretical criteria for theory choice, next to the empirical evidence. But the foregoing suggests that the inferences provided by exploratory factor analysis are not of this latter kind, because they only employ the empirical evidence, leaving the choice of theory underdetermined. And insofar as they do employ theoretical criteria, for instance when deciding over the rotation of the matrix of factor loadings, there is no justification of these criteria that derives from factor analysis itself, or that somehow drops out of the data. The burden of proof is on proponents of a methodological role of factor analysis to provide the justifications for these theoretical criteria.

As indicated in the above in reaction to the problem of underdetermination, Haig states that exploratory factor analysis is only one of several submethods of ATOM. It is only intended to bring down the number of alternative theories to manageable size, after which other submethods can kick in to make a decision on which theory is best. But recall that there are really three separate underdetermination problems at stake, one of

them to do with model selection and two others concerning unidentified models, namely the rotational freedom of factor loadings and the indeterminacy of factor scores. For exploratory factor analysis to bring down the number of possible theories substantially, we may have to get rid of one or two of these problems of underdetermination before entering other submethods. And getting rid of these problems means more than simply providing further theoretical criteria, because in a scientific methodology we would also need a justification of them.

For present purposes we may say that exploratory factor analysis perhaps goes some way towards limiting the number of plausible theories, but that we are still left with an abundance of theories to choose from. Presumably, the theoretical criteria and their justifications needed for making this choice are then provided by the other submethods in ATOM, first by analogical reasoning in the process of theory development, and then by Thagard's theory of explanatory coherence in theory appraisal. It is to an assessment of these two submethods that I now turn.

Everything is Everything: Analogy and the Development of Theory

In this section I focus primarily on the use of analogy in theory development, and argue that as such it leaves theory development largely underspecified. My argument hinges on a mathematical theorem which states that any formal structure can be mapped onto any other formal structure such that truths in the one structure are conserved in the other, provided that both models have the same number of objects. Thus a model is basically analogous to any other model with an equal number of objects, leaving the development of theories by analogy largely unspecified. The conclusion is that as such, analogical

considerations cannot impose any constraints on the theories generated by exploratory factor analysis.

In the submethod of ATOM that deals with theory development, analogical reasoning has two main applications. On the one hand we may expand a theory further into the theoretical realm by analogy to a more fully developed theory. For example, a dynamics systems theory on cognitive development may be given a further theoretical development by analogy to statistical mechanics, adopting certain concepts that have proved useful in the latter theory. On the other hand, we can also use analogical reasoning to prepare a theory for evaluation, working out the empirical consequences of a theory by analogy to a theory that is already testable. Examples of the latter are the formulation of a statistical model of the origin of biblical manuscripts by analogy to statistical models of phylogenetic trees in biology, and the derivation of differential equations for modeling psychological states by analogy to such equations in ecological modeling. Indeed, examples of analogical reasoning in science are abundant, and often analogies are at the heart of creative science.

Recall that after applying the first submethod of ATOM, we are faced with a number of theories that have initial plausibility. The method of analogy is supposed to develop each of these theories, expanding their theoretical and empirical content and possibly discarding a number of them because they sit badly with the analogies exploited. Can analogy indeed perform this function? In the following I develop an argument to the effect that applying analogical reasoning to the theories can lead to indefinitely many alternative theories, all of which will have the same initial plausibility. The problem is that development by analogy is not determinate, because the scientist has multiplicity of

possible analogies to choose from. Unless the analogy is suitably restricted, the application of analogical reasoning will only widen the scope of theories admitted by ATOM.

To run the argument I need a precise definition of an analogy. I will say that an analogy relation can obtain between two models, where a model is a theoretical structure consisting of a set of abstract objects or notions, some operations on objects or notions, and some relations between them. Then one model stands in an analogy relation to another model if there exists an isomorphism between them, that is, if there is a one-to-one function between them such that every relation holding between, and every operation performed on, the objects in the one model corresponds to a relation between, respectively an operation on, the images of these objects, namely the objects in the other model. Note that this is quite a strict form of analogy, because in some cases we may already want to speak of an analogy relation between models if a specific part of the one model structurally resembles a part of another model. In any case, the argument that I develop also holds for this weaker notion of analogy. Another, more serious worry is that the notion of analogy employed here only concerns the formal structure and not the content of the models. I discuss this worry more extensively below.

On the assumption of this formal notion of analogy, I can now make precise the reason that analogical considerations do not lead to determinate theory development. This reason originates in the discussion of a philosophical position called structural realism, defended by Worrall (1989), Ladyman (1998) and Votsis (2005) among many others, according to which scientific theories are about the real world only insofar as they fix the formal structure of their models. The entities that they postulate need not exist, the

structures of their models does. In reaction to one of the early proponents of this view, Bertrand Russell, the mathematician Newman (1928) proved that any collection of things in the world can be organized in such a way that they match this structure, provided that there are the right number of them. Demopoulos and Freedman (1985) point out that this point is similar to Putnam's (1978, p. 123) model-theoretic argument, which shows that as long as the candidate scientific theory matches the observational structures that it is supposed to predict and explain, it can be deemed true in its entirety, hence also on its theoretical part. In brief, the import of this discussion is that the requirement that the world matches the formal structure of a scientific model is almost vacuous.

Newman and Putnam made their points about the relation between a scientific model and the world, which is of course not a formal structure but a physical thing. However, the very same arguments can be run, *a fortiori*, on the relation between two formal models and, in particular, it can be run on the relation between two scientific models that are supposed to be analogous to one another. By Newman's results, we can say that any model is analogous to any other model that has the same number of objects in it. So, by gerrymandering the properties and relations between the objects in the models, we can argue that there is an analogy between a theory on the development of psychological states and a theory on the origin of biblical texts. And both these theories are again analogous to statistical mechanics! It will be clear that the possibilities of variation on this theme are infinite. The conclusion must be that the use of analogy does not present any restrictive guidance in theory development, and that simply following the advice of ATOM to use analogy leaves us with too many possible theoretical developments to choose from.

Clearly, the above argument can hardly be considered a straightforward criticism of analogical reasoning as a submethod of ATOM. It seems more appropriate to say that the notion of analogy employed in the above argument is not true to the intentions of Haig. Accordingly, the real problem with ATOM is that it does not specify the notion of analogy in enough detail, so that for a malevolent philosopher it becomes vulnerable to the above permissiveness. It is notable that this situation is in a sense similar to the underdetermination problems of the preceding section. The submethod of theory generation by means of factor analysis is also not spelled out in sufficient detail to avoid problems of underdetermination. Both in the case of theory generation and theory development, it seems that we are in need of further restrictions, theoretical or empirical, before we can expect the methodology to perform a normative function. At the same time there is no argument to the effect that such restrictions are impossible. One possible source for these restrictions is the third main submethod of theory evaluation, which will be dealt with in the next section.

Resolving Underdetermination: Experiment and Control

The foregoing sections have built a case for the claim that ATOM is too permissive to be a fully fledged methodology of psychological science. This section presents some suggestions to remedy this problem of permissiveness, based on an analytic characterization of experimental practice. The section falls into two parts. First I argue that the theory of explanatory coherence cannot be expected to provide the constraints required to get the other submethods going, and then I examine other ways to remedy the permissiveness of the first two submethods. Specifically, I discuss the use of

experimental interventions to resolve statistical underdetermination, and I consider the use of the experimental setup to fix the reference of key theoretical terms, thus isolating useful analogy relations. These suggestions will be seen to find a natural place in ATOM, as a methodology that covers both discovery and justification.

According to ATOM, the submethods of theory generation and development are followed by the submethod of theory evaluation. At the latter stage, we have a collection of well worked out scientific theories that can all account for the empirical facts, and have some initial plausibility. The proposal of Haig is to then apply the theory of explanatory coherence, as presented by Thagard (2000). This theory measures the rival scientific theories according to their explanatory coherence, made precise in a number of criteria, and then advises to pick the theory that scores best. The theory appraisal part of ATOM is thus a genuine inference to the best explanation, and it is therefore vulnerable to the standard objections to this mode of inference, labeled Hungerford's and Voltaire's objection by Lipton (2004). Hungerford's problem is that the measure of explanatory coherence depends crucially on the exact list of explanatory criteria and the way that these criteria are weighted. There are many ways of making the measure of explanatory coherence precise, hence there will be many different theories that come out best under some such measure. But if ATOM as a whole is supposed to be conducive to finding true scientific theories, there has to be some principled story as to why the specific criteria used in the theory of explanatory coherence further this truth-conduciveness.

For the sake of argument, I will assume that Hungerford's objection is adequately taken care of in the naturalist justifications of Thagard. However, Voltaire's objection remains: how can we be sure that the collection of scientific theories from which we

select the best one contains the true, or at least an approximately true theory? Realist philosophers of science have given a variety of answers to this question. Presumably the answer of ATOM is provided in its other two submethods, which are intended to make sure that the members of the collection of theories all have some initial plausibility. But I have argued in sections 3 and 4 that, as they stand, both these submethods are malfunctioning. I take it that we cannot submit an infinity of theories to an assessment of explanatory coherence, yet we have not been provided with sufficient reason to prefer one finite selection of theories over another. In other words, because of the permissiveness of the other two submethods, the theory of explanatory coherence becomes vulnerable to Voltaire's objection.

Now one possible response to this is to apply the notions at work in explanatory coherence to resolve the permissiveness of the other two submethods. That is, the same criteria that occur in a measure of explanatory coherence can perhaps be used to resolve the underdetermination of exploratory factor analysis and analogical reasoning. But I do not think that this is in line with ATOM itself, and I will not investigate that possibility in the following. Instead I want to suggest how the two problematic submethods of ATOM may be elaborated and improved on by a further examination of the scientific practice to which they apply, specifically of experimental practice. The foregoing may have created the impression that ATOM's occupation with the scientific practice of theory generation and development is responsible for its weakness. Against this, I maintain that this focus on practice also makes for its greatest promise.

As said, the central notion in these suggestions is that of scientific experiment. Consider first the practice of theory generation. Historical studies, for example by

Galison (1987), Steinle (1997), and Chang (2004), indicate that the generation of new concepts and theories is intimately tied up with experimental activity. Viewing theory generation from the perspective of ATOM, I think we can advance a very fitting explanation for this fact. As illustrated in the section on exploratory factor analysis, the generation of theory is strongly underdetermined by the empirical facts. But if we allow for interventions and controls in the world as well as for observations of it, we can greatly enhance the potential of the empirical facts to determine theory. It therefore seems promising to try and extend the technique of exploratory factor analysis with the notions of experimental intervention and control, to thereby resolve the permissiveness of theory generation by explanatory factor analysis.

It leads us too far into the technicalities of factor analysis to make this suggestion precise, but I do want to provide a rough sketch of the idea. A natural starting point for the inclusion of experimental activity is the theory of Bayesian networks. Pearl (2000) and Spirtes et al. (2001) have worked out how such networks, strictly speaking just convenient graphical representations of probability functions, can be used to capture the causal relations in a scientific model, and how the causal networks can be employed for computing expectations over the results of experimental interventions and controls. The idea is to take the insights and formal tools associated with experiments in Bayesian and causal networks, to transport them to the discussion on factor analysis, and to show how the various problems of underdetermination in factor analysis may, at least in part, be resolved by means of these tools. The submethod of ATOM that deals with theory generation can thus be supplemented with detailed advice on what experimental interventions to perform in order to avoid being too permissive.

Next, consider the permissiveness of the second submethod of ATOM, theory development by analogical reasoning. I have already indicated that the permissiveness of analogy considerations is not a criticism of ATOM as such, but rather points to the fact that the notion of analogy is not sufficiently specified in ATOM to do the conceptual work it is supposed to do. Recall that the underdetermination of analogy stems from the fact that in the specification of analogy relations, models were treated as purely formal constructions. What we need is a specification of analogies that are somehow interesting or informative for theory development. As pointed out by Frigg (2002), for a model to successfully represent the world we need to specify not only the formal model but also a physical design, a story in which the world is structured in order to fit the bill of the model. In line with Frigg's views, my suggestion is that for a successful analogy relation we also need a further story cashing out how exactly the models that stand in the analogy relations are supposed to be connected. That is, the use of analogy in ATOM can only lead to normative guidelines when it is supplemented with a story to single out the interesting analogies.

Again, a full specification of these ideas will lead us too far away from the main line of this paper, so I will only provide some rough and ready suggestions showing that experimental practice can again be of use here. In his study of the development of mechanics, van Dyck (2005) argues that Galileo used a series of experiment to fix the meaning of the theoretical concepts in his science of motion. That is, the experiments help to shape what Frigg calls the physical design, the story that gives content to the formally specified models of the theory. In the foregoing I suggested that such stories determine the set of relevant analogies. Following van Dyck, who claims that these

stories are effectively produced in the experimental context, I tentatively conclude that the experimental setting will help to select the relevant analogies. But I admit that this latter conclusion is fairly speculative.

Summing up, the suggestion of the present section is that the specifics of experimental practice may be used to improve the first two submethods of ATOM. Experimental interventions and controls may be used to resolve some of the underdetermination problems in exploratory factor analysis, generating further empirical constraints on theory generation. And experiments may further be used to fix the meaning of the theoretical concepts appearing in the models of the theories, thus restricting the possible analogy relations with other models. But it will be clear that these are just suggestions, and that more research into the details of interventions in factor analysis and concept formation in experiment is needed to arrive at genuine claims.

A Methodology of Scientific Practice

I want to end by relating the above suggestions to the first section of this paper, and to the wider discussion of methodology in the philosophy of science. In the first section I indicated that traditional methodology restricts attention to the domain of theory appraisal. ATOM is a valuable contribution to the philosophical debate just because it opens up to the scientific activities that come before the theory is submitted to testing, to wit, theory generation and theory development. But as I have argued in this paper, it is thereby also at risk of losing its normative status. One possible reaction is to embrace the resulting permissiveness and opt for theoretical pluralism. I have suggested here that another possible reaction is to work out the methodology of theory generation and

development in more detail by reference to experimental practice. I think that this reaction will present a genuine advance of philosophical, and perhaps ultimately of practical, methodology. It will bring analytic philosophy of science closer to science itself, and it will bring scientific practice within reach of rationalization.

Acknowledgements

The author wishes to thank Anne Boomsma and Conor Dolan for helpful discussions on factor analysis, and Brian Haig for stimulating discussions during his 2006 stay at the University of Amsterdam. Support by the Leverhulme Trust is gratefully acknowledged. Send correspondence to: Jan-Willem Romeijn, Faculty of Philosophy, University of Groningen, Oude Boteringestraat 52, 9712 GL, Groningen, The Netherlands. Electronic email may be addressed to j.w.romeijn@rug.nl

References

- Chang, H. (2004). *Inventing temperature: Measurement and scientific progress*. New York: Oxford University Press.
- Demopoulos, W. & Freedman, M. (1985). Bertrand Russell's *The Analysis of Matter*: Its historical context and contemporary interest, *Philosophy of Science* 52, 621–39.
- Dyck, M. van (2005). The paradox of conceptual novelty and Galileo's use of experiments", *Philosophy of Science*, 72, 864–875.
- Fraassen, B. van (1989). *Laws and symmetry*. Oxford: Clarendon.
- Frigg, R. (2002). Models and representation: Why structures are not enough. *Measurement in Physics and Economics Project Discussion Paper Series* (pp.). London: London School of Economics, 42 pages.
- Galison, P. (1987). *How experiments end*. Chicago: University of Chicago Press.
- Haig, B. D. (2005a). An abductive theory of scientific method. *Psychological Methods*, 10, 371–388.
- Haig, B. D. (2005b). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research* 40, 303–329.
- Ladyman, J. (1998). What is structural realism? *Studies in the History and Philosophy of Science*, 29, 409–24.
- Lipton, P. (2004). *Inference to the best explanation* (2nd ed). London: Routledge.
- Maraun, M. D. (1996). Metaphor taken as math: Indeterminacy in the factor analysis model, *Multivariate Behavioral Research* 31, 517–38.

- McDonald, R. P. (1974). The measurement of factor indeterminacy. *Psychometrika*, 39, 203–221.
- Mulaik, S. A. (1985). Exploratory statistics and empiricism. *Philosophy of Science*, 52, 410–430.
- Newman, M. H. A. (1928). Mr Russell's 'Causal theory of perception', *Mind*, 37, 137–148.
- Pearl, J. (2000). *Causality*. Cambridge, MA: MIT Press.
- Psillos, S. (1999). *Scientific realism: How science tracks truth*. London: Routledge.
- Psillos, S. (2002). Simply the best: A case for abduction. In A. C. Kakas & F. Sadri (eds.), *Computational logic* (pp. 83-93). Heidelberg: Springer-Verlag
- Putnam, H. (1978). *Meaning and the moral sciences*. Boston: Routledge & Kegan Paul.
- Putnam, H. (1981). *Reason, truth, and history*. Cambridge: Cambridge University Press.
- Spirtes, P., Scheines, R., & Glymour, C. N. (2001). *Causation, prediction, and search* (2nd ed), Cambridge, MA: MIT Press.
- Steiger, J. H. (1979). Factor indeterminacy in the 1930's and the 1970's: some interesting parallels. *Psychometrika*, 44, 157–167.
- Steinle, F. (1997). Entering new fields: Exploratory uses of experimentation", *Philosophy of Science*, 64, S65-S74.
- Thagard, P. (2000). *Coherence in thought and action*, Cambridge, MA: MIT Press.
- Votsis, I. (2005). The upward path to structural realism. *Philosophy of Science* 72, 1161–1172.
- Worrall, J. (1989). Structural realism: The best of both worlds. *Dialectica*, 43, 99–124.